Jitao Wang

Modern Thermodynamics

Based on the Extended Carnot Theorem

Jitao Wang

Modern Thermodynamics

Based on the Extended Carnot Theorem

With 139 figures

Science Press
Beijing

Springer

Author
Jitao Wang
Microelectronics Department
Fudan University
Shanghai 201203, China
E-mail: jtwang@fudan.ac.cn

ISBN 978-7-03-025565-5
Science Press Beijing

ISBN 978-3-642-11348-2 ISBN 978-3-642-11349-9 (eBook)
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2009941647

Cover design: Frido Steinen-Broo, EStudio Calamar, Spain

Printed on acid-free paper

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

Preface

A senior American Professor, C. Truesdell (1919—2000), said:

> Every physicist knows exactly what the first and second laws mean, but it is my experience that no two physicists agree on them.

Based on my own experience in the last 20 years, Truesdall's experience or feeling was correct. It meant that nobody really understood "What is the second law of thermodynamics?" in the whole 20th century.

It is well known that Carnot theorem is the foundation stone of thermodynamics (especially of classical thermodynamics) and the starting point of the second law of thermodynamics. Carnot summarized daily macroscopic experiences of human beings in 1824. Carnot theorem could not be deduced from any basic law of other disciplines of science. Carnot theorem is that the efficiency of a reversible heat engine is the highest. Carnot theorem emphasized "reversibility" of Carnot cycle, and resulted in the establishment of thermodynamics for reversible processes (i.e. equilibrium thermodynamics) together with thermodynamics for irreversible processes (i.e. nonequilibrium thermodynamics) in classical thermodynamics. However, now it has been found that the general necessary and sufficient condition for the highest conversion efficiency of energy is "nondissipation" of process or system but not "reversibility" of process or system. This is an extended Carnot theorem. A nondissipative cyclical process (such as a nondissipative Carnot cycle) must be reversible, so the extended Carnot theorem has already included the Carnot theorem, but not vice versa. The extended Carnot theorem is an extended foundation stone of thermodynamics (especially of modern thermodynamics) and a new staring point of the second law in the current 21st century. The extended Carnot theorem results in the establishment of a new field of nondissipative thermodynamics (i.e. thermodynamics for nondissipative processes) together with dissipative thermodynamics (i.e. thermodynamics for dissipative processes) in modern thermodynamics.

Thermodynamics is a part of science doing research on energy and energy transformation, and its research objects are the change, development and evolution of macroscopic systems consisting of a large number of particles. It is well-known that the nature of the second law of thermodynamics is the irreversibility of time for the development of macroscopic systems, i.e. "arrow of time". However, Carnot theorem emphasized "reversibility". Such a paradox existed and lasted more than 180 years in the past. This was just the root of puzzlement in thermodynamics. Such a stituation of advocating classicism and excluding modernization is rarely found in other disciplines of science, or we may even say that up to now this is unique.

In 1865, to introduce the cyclical process assumption on the basis of Carnot theorem, Clausius said:

> In order to express analytically the second fundamental theorem in the simplest manner, let us assume that the changes which the body suffers constitute a cyclical process, whereby the body returns finally to its initial condition. $\cdots$

Thus, the Clausius inequality or the entropy increase principle for isolated (or adiabatic) systems was obtained. That was the mathematical expression of the second law of classical thermodynamics. However, the "cyclical process" assumption was not included in Clausius inequality or the entropy increase principle at that time. Now the general mathematical expression of the second law of thermodynamics could be written in its entropy production form for any system, i.e. entropy production of any system is never negative [$\mathrm{d_i}S \geqslant 0$]. Adding the "cyclical process" assumption into his mathematical expression, the mathematical expression of the second law in classical thermodynamics should clearly be written as [$\mathrm{d_i}S_\mathrm{p} = \mathrm{d_i}S \geqslant 0$].

In the same famous paper, Clausius also said:

> The second fundamental theorem, in the form which I have given to it, asserts that all transformations occurring in nature may take place in a certain direction, which I have assumed as positive, by themselves, that is, without compensation; but that in the opposite, and consequently negative direction, they can only take place in such a manner as to be compensated by simultaneous occurring positive transformations.

Clausius' "compensation" is now called thermodynamic coupling. The system including only "positive" spontaneous process(es) "without compensation" is called uncoupling system, and the system including "negative" nonspontaneous process(es) "compensated by simultaneous occurring positive" spontaneous process(es) is called coupling system.

The first half of this citation confirmed the expression of [$\mathrm{d_i}S_\mathrm{p} = \mathrm{d_i}S \geqslant 0$] for classical thermodynamics. That is, only simple uncoupling systems are considered in classical thermodynamics. Meanwhile, the second half of this citation meant that the mathematical expression of the second law in modern thermodynamics for complex coupling systems should be clearly written as [$\mathrm{d_i}S_1 < 0$, $\mathrm{d_i}S_2 > 0$ & $\mathrm{d_i}S \geqslant 0$]. Thermodynamic coupling is the watershed between classical thermodynamics and modern thermodynamics. Here, $\mathrm{d_i}S_1$ and $\mathrm{d_i}S_2$ (or $\mathrm{d_i}S_\mathrm{p}$) are the entropy production of nonspontaneous process(es) and the entropy production of spontaneous process(es), respectively. $\mathrm{d_i}S$ is the entropy production of the system, $\mathrm{d_i}S = (\mathrm{d_i}S_1 + \mathrm{d_i}S_2)$. Of course, during Clausius' time there were no terms of "classical thermodynamics" and "modern thermodynamics", but there were still differences between uncoupling systems "without compensation" and coupling systems with compensation.

Based only on the extended Carnot theorem, on the local equilibrium approximation for description of nonequilibrium states and on mathematical expressions of the second law of thermodynamics mentioned above, a basic complete classification of the thermodynamics discipline can be presented by a building-image, as shown in Fig. 0.1. The whole of the thermodynamics building-image based on the extended Carnot theorem is more general than the classical one based on the Carnot theorem.

More progress in modern thermodynamics can be found in this book and in future work. The whole thermodynamics discipline, including classical thermodynamics and modern thermodynamics, seems to be so simple as said in 1881 by Gibbs:

> One of the principle objects of theoretical research in any department of knowledge is to find the point of view from which the subject appears in its greatest simplicity.

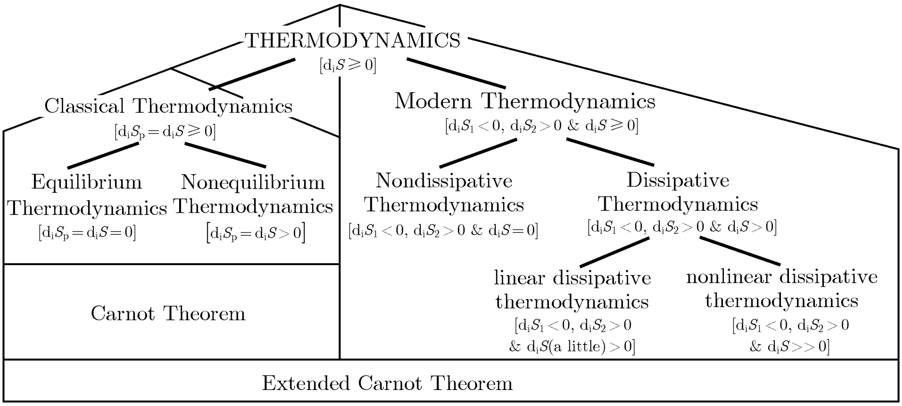


Fig. 0.1 A building-image of the complete thermodynamics discipline

This book can be used as a textbook in universities and colleges for the modernization of the thermodynamics discipline. There are not too much mathematical deductions in this book, so it is also suitable for readers to read and learn by themselves.

Based on the author's experiences, the modernization of the thermodynamics discipline is not easy, so the book is written in three levels. The first level is about 800 words in Preface; the second level is an outline of about 8,000 words in Chapter 1, which was originally a reply in 2006 to a reviewer, a person described in his own words as being "through my personal journey through the realm of equilibrium thermodynamics and nonequilibrium thermodynamics and through my own work therein over decades"; and the third level is the detailed discussions in other parts of the whole book. Thanks to a native British English speaking Ph.D., Dr. David Scott, for his English language corrections, which were made during his stay in Shanghai.

The author is grateful to those organisations which provided financial support in the past: the National Natural Science Foundation of China (NSFC); the Advanced Materials Committee of the High Technology Research and Development Program (863 Program) of China; the Science and Technology Department and the Foreign Affairs Department of the Chinese Education Ministry; and the Fundamental Research Department of the Chinese Ministry of Science and Technology. The author is also grateful to the Swedish Board for Technical Development (Prof. Jan-Otto Carlsson being in charge of the project for my visit to Sweden) for financial support at the very early time.

Fudan University, Shanghai *Prof. Jitao Wang*
May 2010

Contents

Acronyms

ADP	Adenosine DiPhosphate
ATP	Adenosine TriPhosphate
B-Z reaction	Belousov-Zhabotinsky reaction
CALPHAD	CALculation of PHAse Diagrams
CVD	Chemical Vapor Deposition
DNA	DeoxyriboNucleic Acid
FKN mechanism	Field-Körös-Noyes mechanism
FT	Fluctuation Theorem
HFCVD	Hot Filament Chemical Vapor Deposition
HGCTR	High Gradient Chemical Transformation Reaction
HP	High Pressure
HPHT	High Pressure High Temperature
LP	Low Pressure
MPCVD	Microwave Plasma Chemical Vapor Deposition
PCR	Polymerase Chain Reaction
RNA	RiboNucleic Acid
SAH	Superequilibrium Atomic Hydrogen
SAO	Superequilibrium Atomic Oxygen

Chapter 1

An Outline of Modern Thermodynamics

Abstract Thermodynamics is a core part of science. Nearly all scientists should have a basic knowledge of thermodynamics. Thermodynamics is a science of development, and is a viewpoint on scientific development in natural sciences. The achievement of thermodynamics has influence not only on natural sciences, but also on social sciences and philosophy. Fundamental concepts and definitions are very important for any discipline of science, so what thermodynamics is and what the second law of thermodynamics is have become the key points of puzzlement in the thermodynamics discipline. It has been found that the root of puzzlement is the Carnot theorem. The Carnot theorem emphasized "reversibility", but the nature of the second law of thermodynamics is "irreversibility", i.e. "arrow of time". The extended Carnot theorem emphasizes "nondissipation" and results in a new field of nondissipative thermodynamics, which belongs to the equality part of the second law of thermodynamics as well. Some fundamental concepts of thermodynamics are drastically changed. [1]

1.1 Challenges to the Second Law of Thermodynamics Coming from Two Sides

At present, challenges to thermodynamics are coming from two sides. What is "thermodynamics" and what is "the second law of thermodynamics" may become the key points of puzzlement in thermodynamics.

Truesdell (1919–2000) [2] was an American senior professor who wrote two books related to thermodynamics, "*The Tragicomical History of Thermodynamics*, 1822—1854 (Springer, 1980)" and "*Rational Thermodynamics* (Springer, 1984)".[1] He said:

> Every physicist knows exactly what the first and second laws mean, but it is my experience that no two physicists agree on them.

At first glance, his words seemed to be inconsistent, but his experience might give him such a strong feeling. If such an experience or feeling was correct, it meant that the recognition of the second law of thermodynamics in the 20th century was puzzled and has continued to be so. At the beginning of the 21st century, challenges to the second law of thermodynamics are coming from two sides.

The book, "*Challenges to the Second Law of Thermodynamics: Theory and Experi-*

---

[1] http://en.wikipedia.org/wiki/Clifford_Truesdell. Cited 12 Oct 2008

ment”, [2] written by Vladislav Čápek and Daniel Peter Sheehan and D.P. Sheehan in 2005 is a representation of challenges from one side. That book emphasizes that the second law

> Once established, it settled in and multiplied wantonly; ⋯. Not all formulations are equivalent, such that to satisfy one is not necessary to satisfy another. Some versions overlap, while others appear to be entirely distinct laws.

Thus, Cápek and Sheehan listed 21 kinds of the so-called “second law of thermodynamics” and the “expressions of entropy”. Alongside the statements and expressions proposed by Sadi Carnot, Lord Kelvin, Rudolf Clausius, Josiah Willard Gibbs, et al., they also confusedly listed Boltzmann “probability entropy”, Shannon “information entropy”, “entropy” defined by density functions of quantum mechanics and so on. Their list also included Clausius’ “fundamental laws of the universe”. Based on these, Cápek and Sheehan proposed their “quantum challenges” and “gravitational challenges”. Of course, these so-called “the second law of thermodynamics” and “entropy” extended by them are different from those accepted by most thermodynamicists. At least their “challenges” do not belong to the academic discipline of thermodynamics, but belong to “other kind of disciplines” out of thermodynamics.

On the other hand, challenges are coming from a few classical (traditional) thermodynamicists. They mainly aggrandize the role and applicability of classical thermodynamics, and want to limit applicability of thermodynamics to the same as that of classical thermodynamics, i.e. they regard classical (or traditional) thermodynamics as the whole of thermodynamics. For instance, “*Non-equilibrium Thermodynamics and the Production of Entropy —— Life, Earth, and Beyond*”, [3] edited by A. Kleidon and R.D. Lorenz in 2005, is a book in which all kinds of complex systems are intended to be treated by the “maximum entropy principle” on the basis of classical thermodynamics. In their preface they cited Einstein’s famous words in 1949,

> A theory is more impressive the greater the simplicity of its premises is, the more different kinds of things it relates, and the more extended its area of applicability. Therefore the deep impression which classical thermodynamics made upon me. It is the only physical theory of universal content concerning which I am convinced that, within the framework of the applicability of its basic concepts, it will never be overthrown.

In fact, Einstein only said that “within the framework of the applicability of its basic concepts, it (classical thermodynamics) will never be overthrown”, but out of “the framework of the applicability of its basic concepts”, classical thermodynamics may not be suitable anymore.

A new Chinese book on thermodynamics written by Zong-Han Lin in 2007 [4] may be taken as another example, in which the English name of thermodynamics, including “dynamics”, was regarded as “morbidity (or unsuitable)”. His reason was that there is “no time variable included in the theory of traditional thermodynamics”, and he limited thermodynamics into the category of classical thermodynamics. Lin denied the objective existence of “classical thermodynamics” and “nonclassical thermodynamics” (i.e. “modern thermodynamics”) by the quillet that thermodynamics is a macroscopic theory without partition of “classical” and “quantum”. In Lin’s book the usual classical thermodynamics was called “traditional thermodynamics”. That may be OK. However, in that book once the name of “traditional thermodynamics” had been changed into “(traditional) thermodynamics” with a pair of brackets, and once the brackets together with the word of “traditional” had further been deleted, “traditional thermodynamics” was immediately changed into the full discipline of “thermodynamics”. That is a distortion and may leave harmful impressions for young students. The macroscopic theory of thermodynamics may also be divided into a “classical (or traditional)” part and a “modern” part. This is the development of the thermodynamics discipline.

To sum up, what is “thermodynamics” and what is “the second law of thermodynamics” become the key points of puzzlement.

1.2 Root of Puzzlement: Carnot Theorem

Carnot theorem was the foundation stone of thermodynamics and classical thermodynamics. It was also the starting point of the second law of thermodynamics. The extended Carnot theorem is an extended foundation stone of thermodynamics and a new starting point of the second law in the current 21st century. The “reversibility” is not a general necessary condition for the highest conversion efficiency of energy, but “nondissipation” is really the necessary and sufficient condition for the highest conversion efficiency of energy.

It is well known that Carnot theorem and the basic laws of thermodynamics were established on the basis of a large number of human daily macroscopic experiences, but not deduced from any basic law of other disciplines of science. Carnot theorem was the foundation stone of thermodynamics and the starting point of the second law of thermodynamics. Carnot theorem told us: the reversible heat engine is of the highest conversion efficiency of energy among all other heat engines working between the same hotter heat reservoir and the same colder heat reservoir. That was an important human experience at that time (1824) summarized by Carnot. Now that more than 180 years has passed, the human experience has developed, and the framework and the deepness of applicability of thermodynamics are also in development. The research objects of thermodynamics have been extended from heat-work conversion into any kind of energy-conversion under thermodynamic consideration.[5] For instance, the chemical energy of the reaction between methane (CH_4) and oxygen (O_2) in fuel-battery can be converted into electric energy. It can be imagined that if the chemical energy can completely convert into the electric energy without dissipation (i.e. 100 %), then its conversion efficiency must be the highest. Therefore, the efficiency of any nondissipative process or system for macroscopic conversion of energy must be the highest.[6, 7] That is the extended Carnot theorem. Should we pursue a reversible fuel-battery for getting the highest conversion efficiency of energy? No, it is not necessary. Obviously, from the produced carbon dioxide (CO_2) and water (H_2O), and the electric energy released by the fuel-battery, it is impossible to obtain the equal amount of methane and oxygen consumed, because during the synthesis of methane a series of organic hydrocarbon like methane will inevitably be obtained. The “reversibility” in Carnot theorem is not a general necessary condition, but “nondissipation” is just the necessary and sufficient condition for the highest macroscopic conversion efficiency of energy.

Any nondissipative cyclical process must go back to its initial state, so nondissipative Carnot cycle must be reversible. Therefore, the extended Carnot theorem includes the

Carnot theorem, but not vice versa. The extended Carnot theorem is the extended foundation stone of thermodynamics and a new starting point of the second law in the current 21st century.

The nature of the second law of thermodynamics is the irreversibility of time in the development of macroscopic systems, i.e. the characteristics of “arrow of time”. However, Carnot theorem did just emphasize “reversibility”. Such a paradox existed always in classical thermodynamics, and that is what is just the root of puzzlement in thermodynamics and the limitation of classical thermodynamics. Such a status lasted more than 180 years and has rarely been found during the modernization of other disciplines of science, or we may even say that up to now this is unique.

1.3 Uncertainty or Incompleteness of Clausius Inequality

If we first talk about the written statement of what “thermodynamics” is and what “the second law of thermodynamics” is, then the argument will be long lasting because the detailed written statements are at least a little different from each other in different textbooks on thermodynamics. In textbooks of physical disciplines, discussions on mathematical expressions are usually preferred. Therefore, we will first discuss the mathematical expression of the second law of thermodynamics, which is the core of thermodynamics.

Based on the development of thermodynamics in the 19th century and at the beginning of the 20th century, the following three kinds of mathematical expressions of the second law of thermodynamics for different systems are widely accepted.

(1) Entropy increase principle for isolated (or adiabatic) macroscopic systems.

$$(\mathrm{d}S)_{\mathrm{iso}} \geqslant 0. \tag{1.1}$$

Here, $(\mathrm{d}S)_{\mathrm{iso}}$ is the entropy change of an isolated (or adiabatic) system.
leftskip5pt

(2) Gibbs free energy decrease principle for isothermal isobaric macroscopic systems.

$$(\mathrm{d}G)_{T,p} \leqslant 0. \tag{1.2}$$

Here, $(\mathrm{d}G)_{T,p}$ is the Gibbs free energy change of an isothermal isobaric system.

(3) Positive entropy production principle for any macroscopic systems.

$$\mathrm{d_i}S \geqslant 0. \tag{1.3}$$

Here, the entropy change of system $\mathrm{d}S$ has been divided into two parts: The part of entropy production of the system $\mathrm{d_i}S$ (the part of entropy change due to the irreversible processes inside the system), and the other part of entropy flow of the system denoted by $\mathrm{d_e}S$ (the part of entropy change due to exchanges of energy and/or matter with the surroundings), i.e. $\mathrm{d}S = \mathrm{d_i}S + \mathrm{d_e}S$. Under isothermal isobaric conditions, there is a relation between the entropy production and the Gibbs free energy change, i.e.

$$(\mathrm{d}G)_{T,p} = -T\mathrm{d_i}S. \tag{1.4}$$

> Because the positive entropy production principle is suitable not only for isolated systems, closed systems and open systems, but also for isothermal isobaric systems and nonisothermal nonisobaric systems, it is called the general mathematical expression of the second law of thermodynamics. Another advantage of the positive entropy production principle is that the entropy production of a system is directly connected with the internal irreversible processes of the system, so the entropy production of a system is the sum of entropy productions of internal irreversible processes. Therefore, hereafter in this book our discussions are mainly based on the form of entropy production. If someone is not familiar with or is doubtful of the form of entropy production, please convert it into the form of Gibbs free energy under isothermal isobaric conditions by using Eq. (1.4), and it will be more easy to understand.

Note that temperature was defined and measured under heat equilibrium conditions, and the local equilibrium approximation had already been introduced since Carnot and Clausius discussed body with different temperatures. Here, “any macroscopic systems” should also satisfy the requirement of the local equilibrium approximation. If discussions are concerned with nonequilibrium systems, the local equilibrium approximation has already been included. Of course, the mathematical expression of the second law of thermodynamics could also be written in the form of the free energy decrease form: $(\mathrm{d}F)_{T,V} \leqslant 0$ and so on (but do not include Boltzmann “probability entropy”, Shannon “information entropy”, “entropy” defined by density functions of quantum mechanics, and others). We will not discuss them any more here.

It is well-known that the first mathematical expression of the second law of thermodynamics was proposed by Clausius in his famous paper published in 1865. He said: (p. 328-329)[8]

> In order to express analytically the second fundamental theorem in the simplest manner, let us assume that the changes which the body suffers constitute a cyclical process, whereby the body returns finally to its initial condition. ···

In this way he obtained Clausius inequality

$$\mathrm{d}S \geqslant \frac{\delta Q}{T} \tag{1.5}$$

and the entropy increase principle for isolated (or adiabatic) systems, as shown in Eq. (1.1). However, the assumption of “cyclical process” was not included in the Clausius inequality and the entropy increase principle. Based on such a cyclical-process assumption, the entropy change for an arbitrary irreversible process would be calculated on the basis of the entropy change for a reversible process with the same initial equilibrium state and the same final equilibrium state. That is a severe limitation of classical thermodynamics. Both the same initial state and the same final state must be in equilibrium, because in a reversible process the system must be always in equilibrium state. The entropy change of process must be always equal to the entropy change of the system in classical thermodynamics. Therefore, Eq. (1.1), Eq. (1.2) and Eq. (1.3) should be changed a little, as shown in Eq. (1.6), Eq. (1.7) and Eq. (1.8) for classical thermodynamics.

$$(\mathrm{d}S_{\mathrm{p}})_{\mathrm{iso}} = (\mathrm{d}S)_{\mathrm{iso}} \geqslant 0. \tag{1.6}$$

Here, $(\mathrm{d}S_{\mathrm{p}})_{\mathrm{iso}}$ is the entropy change of a process in an isolated (or adiabatic) system.

$$(\mathrm{d}G_{\mathrm{p}})_{T,p} = (\mathrm{d}G)_{T,p} \leqslant 0. \tag{1.7}$$

Here, $(\mathrm{d}G_{\mathrm{p}})_{T,p}$ is the Gibbs free energy change of a process in an isothermal isobaric system.

$$\mathrm{d_i}S_{\mathrm{p}} = \mathrm{d_i}S \geqslant 0. \tag{1.8}$$

Here, $\mathrm{d_i}S_{\mathrm{p}}$ is the entropy production of a process.

In the same paper, Clausius also said: (p. 364)[8]

> The second fundamental theorem, in the form which I have given to it, asserts that all transformations occurring in nature may take place in a certain direction, which I have assumed as positive, by themselves, that is, without compensation; but that in the opposite, and consequently negative direction, they can only take place in such a manner as to be compensated by simultaneous occurring positive transformations.

In the first half of this citation, it is clearly confirmed that only the spontaneous process(es) in "positive" direction and "without compensation" is(are) under consideration in classical thermodynamics, so the Eq. (1.8) is correct for classical thermodynamics. Now the "compensation" is called thermodynamic coupling, so systems including only spontaneous process(es) in "positive" direction and "without compensation" are called simple uncoupling systems or uncoupling systems for short. That is, Eq. (1.8) is the mathematical expression of the second law of thermodynamics for simple systems (or better to say, for uncoupling systems) in classical thermodynamics, and the cyclical-process assumption in classical thermodynamics has already been included in Eq. (1.8). Once it is found that the entropy production of an internal process does not agree with the entropy production of the system, classical thermodynamicists like to call the corresponding process "abnormal", "mystery", or maybe "violating the second law of thermodynamics" and so on.

In the second half of this citation, it is clearly confirmed that both the spontaneous process(es) in "positive" direction and the nonspontaneous process(es) in "negative" direction with compensation are under consideration in modern thermodynamics. Therefore, the mathematical expression of the second law of thermodynamics for complex systems (or better to say, for coupling systems) in modern thermodynamics should be more clearly written as

$$\mathrm{d_i}S_1 < 0, \mathrm{d_i}S_2 > 0 \ \& \ \mathrm{d_i}S \geqslant 0. \tag{1.9}$$

The entropy production of nonspontaneous process(es) in the system is negative, i.e. $\mathrm{d_i}S_1 < 0$; the entropy production of the spontaneous process(es) is positive, i.e. $\mathrm{d_i}S_2 > 0$; and after compensation the entropy production of the system is never negative, i.e. $\mathrm{d_i}S = (\mathrm{d_i}S_1 + \mathrm{d_i}S_2) \geqslant 0$.

Clausius did not write out the mathematical expression of the second law of thermodynamics for complex systems with compensation for modern thermodynamics. Of course, the terms, "classical thermodynamics" and "modern thermodynamics" did not exist at that time, but Clausius had already divided systems into two kinds of systems: systems "without compensation" and "compensated" systems, i.e. uncoupling systems and coupling systems. That is clear enough to know that thermodynamic coupling, or called "compensation”, is the watershed between classical thermodynamics and modern thermodynamics, and that thermodynamic coupling is the core of modern thermodynamics.

To sum up, the uncertainty or incompleteness of Clausius mathematical expressions at that time was: (i) Clausius inequality did not include his cyclical-process assumption; (ii) Clausius did not write out his “compensation” in mathematical expression. Now mathematical expressions of the second law of thermodynamics should be written as $[\mathrm{d_i}S_\mathrm{p} = \mathrm{d_i}S \geqslant 0]$ for simple uncoupling systems in classical thermodynamics; and for complex coupling systems in modern thermodynamics it should be written as $[\mathrm{d_i}S_1 < 0, \mathrm{d_i}S_2 > 0 \ \&\ \mathrm{d_i}S \geqslant 0]$.

1.4 Classification of Thermodynamics in the Current 21st Century

From the extended Carnot theorem and the correct understanding of the second law of thermodynamics in the current 21st century, a basic and complete classification of thermodynamics can be directly and exactly constructed, as shown in Fig. 1.1.[9~12] The corresponding mathematical expressions of the second law of thermodynamics for each field (or subfield) and their individual frameworks of applicability can be more exactly listed in Table 1.1.

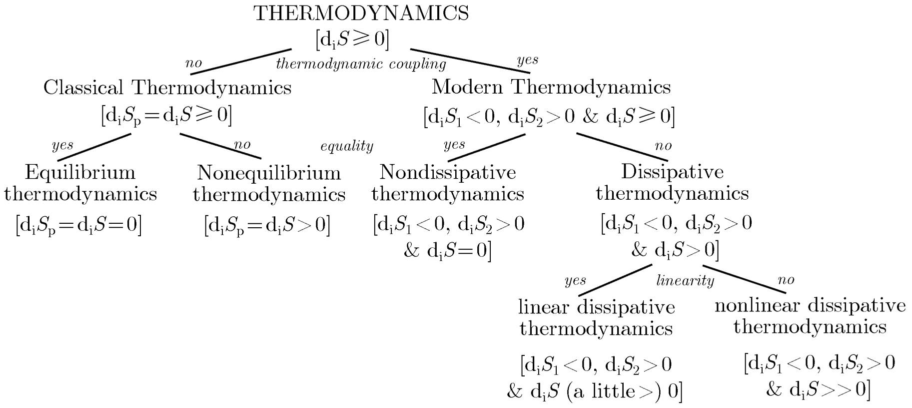


Fig. 1.1 A basic and complete classification of thermodynamics in the current 21st century. $\mathrm{d_i}S_\mathrm{p}$, $\mathrm{d_i}S_1$ and $\mathrm{d_i}S_2$ are entropy productions of process(es). $\mathrm{d_i}S$ is entropy production of the system, $\mathrm{d_i}S = (\mathrm{d_i}S_1 + \mathrm{d_i}S_2)$.

It should be pointed out that there were at least two books published in 1933[13] and in 1998[14], respectively, with the same book title of “*Modern Thermodynamics*”. The subtitle of the former was “*by the methods of Willard Gibbs*”, and the subtitle of the latter was “*from heat engines to dissipative structures*”. Of course, in both these books “modern thermodynamics” meant the current status of thermodynamics at that time. In this book the current status of thermodynamics is still named “thermodynamics”, while the meaning of “modern thermodynamics” here is mainly defined and used as an antonym of “classical thermodynamics”. The title of this book, “Modern thermodynamics”, is an exception as it has a little extension to mean the whole modernized thermodynamics discipline.

Table 1.1 Mathematical expressions of the second law of thermodynamics with applicability frameworks for each field

Name of field	Mathematical expression of the second law
Thermodynamics	[$\mathrm{d_i}S \geqslant 0$] for any macroscopic systems
Classical thermodynamics	[$\mathrm{d_i}S_\mathrm{p} = \mathrm{d_i}S \geqslant 0$] for simple (uncoupling) systems
Equilibrium thermodynamics	[$\mathrm{d_i}S_\mathrm{p} = \mathrm{d_i}S = 0$] for equilibrium systems
Nonequilibrium thermodynamics	[$\mathrm{d_i}S_\mathrm{p} = \mathrm{d_i}S > 0$] for simple (uncoupling) nonequilibrium systems
Modern thermodynamics	[$\mathrm{d_i}S_1 < 0$, $\mathrm{d_i}S_2 > 0$ & $\mathrm{d_i}S \geqslant 0$] for complex (coupling) systems
Nondissipative thermodynamics	[$\mathrm{d_i}S_1 < 0$, $\mathrm{d_i}S_2 > 0$ & $\mathrm{d_i}S = 0$] for nondissipative systems
Dissipative thermodynamics	[$\mathrm{d_i}S_1 < 0$, $\mathrm{d_i}S_2 > 0$ & $\mathrm{d_i}S > 0$] for dissipative systems
Linear dissipative thermodynamics	[$\mathrm{d_i}S_1 < 0$, $\mathrm{d_i}S_2 > 0$ & $\mathrm{d_i}S$ (a little)$>$ 0] for linear dissipative systems
Nonliear dissipative thermodynamics	[$\mathrm{d_i}S_1 < 0$, $\mathrm{d_i}S_2 > 0$ & $\mathrm{d_i}S >> 0$] for nonlinear dissipative systems

As shown in Fig. 1.1 and Table 1.1, the basic and complete classification of thermodynamics is obtained only on the basis of the second law of thermodynamics. Therefore, the simple out-of-date classifications based merely on the equality and inequality to divide thermodynamics into several pairs of "reversible" and "irreversible", "equilibrium" and "nonequilibrium", or "equilibrium state" and "nonequilibrium state" should be avoided.

The second law of thermodynamics in the current 21st century here includes mainly the contents incarnated by the mathematical expressions of the second law of thermodynamics for each field, the extended Carnot theorem, and the following discussed dissipation decrease theorem. The connotations as shown in Fig. 1.1 and Table 1.1 are a step-up and are quite different from the contents of thermodynamics textbooks in the 19th century and in the 20th century.

1.5 A Typical Case of Nondissipative Thermodynamics: Nonequilibrium Phase Diagrams

Classical thermodynamics considered that the entropy production of the transformation from graphite to diamond under low pressure was negative, so it was regarded as thermodynamically impossible to realize. Modern thermodynamics considers that the entropy production of reaction (1) (i.e. the transformation from graphite to diamond under low pressure) is negative, so it is impossible to take place alone. But under the drive of reaction (2) (i.e. the association of superequilibrium atomic hydrogen with positive entropy production), reaction (1) is possible to take place simultaneously. That is the thermodynamic coupling model for activated low-pressure diamond growth.

From 1970 up to now, the activated low-pressure diamond growth has developed from film or polycrystalline diamond growth to about 10 carat gem-quality single-crystal diamond growth. However, as we know, the corresponding reasonable thermodynamic explanation is still provided only on the basis of our thermodynamic coupling model for activated low-pressure diamond growth [15~17], and the corresponding calculation of nonequilibrium phase diagram can be done only based on nondissipative thermodynamics as well.

In around 1970 the new technology of activated low-pressure diamond growth was achieved in the former Soviet Union. However, due to the influence of equilibrium phase diagram of classical thermodynamics, the new technology was joked as "alchemy" by some scientists for about 16 years. In about 1980, this technology was also achieved in Japan, and this Japanese laboratory was visited by an American scientist in 1984. In 1986 the American scientist repeated the experiment and announced the achievement to news media. The new diamond technology suddenly became a hot point of research projects around the whole world. In February 1990, a review paper [18] published in the journal of *Science* said:

> All of these formulations suffer from being inconsistent with at least some experimental results.

Note that "All of these formulations" mentioned in this citation must have been published at that time.

In April 1990, the author, as the first speaker at *the Diamond Symposium of the 8th International Conference on Thin Films* proposed a thermodynamic coupling model for the activated low-pressure diamond growth from the vapor phase (J.-T. Wang and J.-O. Carlsson were coauthors of this paper). [19] The model was called thermochemical model or chemical pump model at that time. In 1995, the author further calculated nonequilibrium phase diagrams for the activated low-pressure diamond growth. [20] Later a series of nonequilibrium phase diagrams were calculated by the author's group. Theoretical results agree very well with the world-wide reported reliable experimental data. In 2002, a new field of nondissipative thermodynamics was creatively established by the author. [17] Some reviews published in journals or books provided evaluations of this progress.

For instance, in 2001 Y. Gogotsi, S. Welz, D.A. Ersoy et al. in their article published in *Nature* said: [21]

> ··· the thermodynamic coupling model that can be used to explain the formation of diamond in reaction (1) and (2) have been well developed.

This concerned our paper published in the journal of American Electrochemical Society in 1994 by the author and coworkers. [22]

For nonequilibrium phase diagrams of the activated low-pressure diamond growth, B. Heimann, V. Raiko and V. Buck made a citation of a series of our 5 papers on nonequilibrium phase diagrams and evaluated, [23]

> The projection of phase diagrams obtained from the thermodynamic coupling model may be crucial for understanding Bachmann's (empirical phase) diagram.

In 2002, an English edition book called "*Nonequilibrium Nondissipative Thermodynamics*" written by the author was published by Springer. [17] It was a symbol of the establishment of a new field of nondissipative thermodynamics. In 2004, there were two reviews of the book published in Europe and the US, respectively. M. Pons and D. Ballutaud said at the end of their book review, [24]

> Finally, the book is an excellent bibliographic source, for the fields of both thermodynamics and diamond growth. It is certainly worth reading and the concepts presented worth thinking about.

J.D. Felske said at the end of another book review, [25]

> Overall, this book should be of interest not only to those who work in diamond film production but to thermodynamicists and biochemists as well.

In China, an evaluation in 2000 also said, [26]

> There were some creations in China on the researches of diamond films, such as 'thermodynamic coupling model of the activated low-pressure diamond growth' and the entirely new concept of 'phase diagrams of nonequilibrium stationary states' were proposed firstly in the international forums. All of these progresses were focused attention upon.

In 2001, a book on low-pressure diamonds written by D.-H. Dai and K.-S. Zhou clearly pointed out, [27]

> Comparing with Bachmann's empirical C-H-O ternary phase diagram, $\cdots$ the calculated C-H-O ternary phase diagrams of nonequilibrium stationary states on the basis of nonequilibrium thermodynamics are much better in the coverage of experimental data and academic significances. Authors believe that the nonequilibrium phase diagrams are more perfect and their instructions on the technology are of more strong persuasions.

and so on. In 2004, there were also some similar evaluations in the publication by D.-H. Dai, K.-S. Zhou, Z.-H. Yuan et al. [28] The author will not repeat them here.

1.6 A Typical Case of Dissipative Thermodynamics: Spiral Reactions

Classical thermodynamics had considered that if the entropy production of the forward reaction was positive, then the entropy production of the backward reaction would surely be negative, so chemical oscillation was regarded as impossible in classical thermodynamics. Based on modern thermodynamics in the current 21st century it is known that the nature of chemical oscillations is a performance of spiral reactions. [6,10,29~31] Chemical oscillations are side views of spiral reactions. Both entropy productions of “the forward reaction” and “the backward reaction” in such a chemical oscillation are positive, so chemical oscillations can really take place. The planform of the spiral reaction is a cyclical reaction. Any practically progressing cyclical reaction must be dissipative, i.e. with a decrease of Gibbs free energy under isothermal isobaric conditions or with a positive entropy production in any cases. Nondissipative cyclical reaction is an ideal limitation of the second law of thermodynamics. It will not take place practically, but it does not violate the second law of thermodynamics.

In around 1958, the phenomenon of chemical oscillations, which was found and confirmed many times, had much puzzled classical thermodynamicists. Among these chemical oscillations, Belousov-Zhabotinsky reactions (or called B-Z reaction for short) are representative. The kinetic mechanism of these reactions is now very clear. The famous and recognized FKN mechanism was commonly proposed by three scientists (Field, Körös and Noyes). They pointed out that after each period (or cycle) of oscillation a general reaction equation has been finished. That is, a certain amount of organic acid has been oxidized into carbon dioxide and water. Details of the reaction are very complex. Usually, it can be divided into more than 20 reaction steps. Of course, some of them may be combined with each other. For instance, the cyclical reaction might be divided into three steps of reaction A, reaction B and reaction C. Due to the escape of carbon dioxide bubbles from the solution and the formation of water originally being in a large amount in the solution, after a period of the cycle the reaction seems to be repeated at the beginning of a new cycle of reactions A, B and C.

On the basis of modern thermodynamics in the current 21st century it has been found that B-Z reactions are reactions with the spiral accumulation of entropy production (or called spiral reactions for short, i.e. the sum of entropy production of the system seems to accumulate along a spiral stairway). [6, 10] B-Z reactions are the result of the following causes:

1. Reaction A, reaction B and reaction C all as a whole reaction of the system at different time intervals can go forward, so all of them must accord with the second law of thermodynamics, and all of their entropy productions must be positive, i.e. $\Delta_i S_A > 0$, $\Delta_i S_B > 0$ and $\Delta_i S_C > 0$.
2. Oxidation of organic acid is a very strong energy-supply reaction. The entropy production of the total reaction within a period of oscillation (i.e. W = A + B + C) must be also positive, i.e. $\Delta_i S_W > 0$.
3. Reaction A, reaction B and reaction C are all thermodynamic coupling reactions, i.e. $[\Delta_i S_{[(-B)+(-C)]} < 0, \Delta_i S_W > 0$ & $\Delta_i S_A > 0]$, $[\Delta_i S_{[(-C)+(-A)]} < 0, \Delta_i S_W > 0$ & $\Delta_i S_B > 0]$ and $[\Delta_i S_{[(-A)+(-B)]} < 0, \Delta_i S_W > 0$ & $\Delta_i S_C > 0]$. Here, $A = W + [(-B) + (-C)]$, $B = W + [(-C) + (-A)]$ and $C = W + [(-A) + (-B)]$. Therefore, at the same time of each reaction step, the total oxidation reaction and other backward reactions take place simultaneously. In this way, cyclical reactions are continuously going on.

That is, both the chemical oscillations and cyclical reactions can be well understood in the applicability framework of modern thermodynamics. Side views of a spiral reaction are chemical oscillations, while planform of a spiral reaction is a cyclical reaction. Any practically occurring cyclical reaction must be dissipative, i.e. with a decrease of Gibbs free energy under isothermal isobaric conditions or with a positive entropy production in any cases. Nondissipative single-way cyclical reactions are limitations of the second law of thermodynamics. Nondissipative single-way cyclical reaction will not practically take place, but it does not violate the second law of thermodynamics.

1.7 Out-of-Thermodynamics Assumption for Reciprocal Relations

Onsager reciprocal relations were positioned in the inequality part of the second law of thermodynamics. Just as said by Onsager himself, the derivation of his reciprocal relations were on the basis of the introduction of an out-of-thermodynamics "assumption of microscopic reversibility" or assumption of "detailed balancing". Therefore, they are not strict thermodynamic quantitative relations. Experimental data also showed their approximation.

As Onsager said, the reciprocal relations were first deduced by Thomson (Kelvin), Helmholtz and others, but their deductions were not complete or based on special (such as thermoelectrics or electrolysis cell) experiments, so Onsager called them quasi-thermodynamics.

Onsager also clearly pointed out,[32, 33]

> a general class of reciprocal relations in irreversible processes will be derived from the assumption of microscopic reversibility.

For instance, his reciprocal relations could be derived in a cyclical reaction with an "assumption of microscopic reversibility" or the assumption of "detailed balancing", but these assumptions are not thermodynamic theorems or laws. Therefore, Onsager reciprocal relations actually still belonged to quasi-thermodynamics. We should never absolutely use it as a fundamental thermodynamic law. On a website, Onsager reciprocal relations have been called the fourth law of thermodynamics.[34] This is not correct. Experimental data, which sometimes could be only 77 % of the theoretical value,[14] had also confirmed the approximation of Onsager reciprocal relations.

In the first sentence of the abstract of Onsager's first famous paper in 1931, Onsager wrote:[32]

> Examples of coupled irreversible processes ··· are discussed.

The nature of reciprocal relations is the relation between coupling coefficients of different irreversible processes, or in common words, the reciprocal relations are corresponding interactions between irreversible processes, i.e. the influence of irreversible process i on irreversible process j should be equal to the influence of irreversible process j on irreversible process i. Such a problem is very easy to treat in thermodynamic coupling model. That is, if there is no dissipation, the absolute values of two entropy productions are equal but their signs are opposite. Due to the historical limitations, Onsager had to introduce an out-of-thermodynamics assumption and empirical linear phenomenological relations for deducing the approximate equality of reciprocal relations between coupling coefficients. Now we know that the single-way cyclical reaction is an ideal limitation of spiral reactions.

1.8 Out-of-Thermodynamics Model for Dissipative Structures

Prigogine dissipative structures were established on the basis of nonlinear positive-feedback kinetic "Brusselator", i.e. an out-of-thermodynamics model. Besides, Prigogine agreed with the expression of thermodynamic coupling, but he misunderstood it during its applications.

Chemical oscillation was an important example and application of Prigogine dissipative structures. Prigogine also clearly pointed out that the theory of dissipative structures was established on the basis of an out-of thermodynamics nonlinear positive feedback kinetic model, called "Brusselator". Therefore, it is easy to understand why some thermodynamics textbooks often omitted Prigogine dissipative structures, and allowed the theory of dissipative structures to be out of the contents of thermodynamics courses.

Now it has been found that the nature of both cyclical reactions and chemical oscillations in nature is a performance of spiral reactions. Spiral reactions were discovered directly on the basis of the mathematical expression of thermodynamic coupling, i.e. the mathematical expression of the second law in modern thermodynamics. Onsager reciprocal relations and Prigogine dissipative structures usually belong to linear dissipative thermodynamics and nonlinear dissipative thermodynamics, respectively. There seems to be no substantial difference between these two fields, but in the modern classification of thermodynamics (Fig. 1.1 and Table 1.1) they are still preserved. Care should be taken that the terms of “close-to-equilibrium” (or “nearly in equilibrium”) and “far-from-equilibrium” should not be used to substitute the terms of “linear” and “nonlinear” in modern thermodynamics. This is because “close-to-equilibrium” (or “nearly in equilibrium)” means that the state function is close to an extremum, but the state function in the interesting region under nonequilibrium dissipative conditions often has no extremum.

It should be pointed out that Prigogine agreed with the concept of thermodynamic coupling and its mathematical expression, but he misunderstood the expression of thermodynamic coupling in its applications by adding “under standard conditions”.[35] That would be the cause for him to incorrectly consider an ordinary chain reaction of HBr formation as an example of thermodynamic coupling.[14]

1.9 Dissipation (or Entropy Production) Decrease Theorem

Dissipation (or entropy production) decrease theorem is also generally applicable. From the mathematical expression of thermodynamic coupling [$\mathrm{d_i}S_1 < 0$, $\mathrm{d_i}S_2 > 0$ & $\mathrm{d_i}S \geqslant 0$], it can be found that if the entropy production of the system agrees with $\mathrm{d_i}S \geqslant 0$, then spontaneous process, i.e. positive entropy production process ($\mathrm{d_i}S_2 > 0$), is a kind of potential driving force to drive any possible nonspontaneous process, i.e. negative entropy production process ($\mathrm{d_i}S_1 < 0$). For simple uncoupling systems, there is no possibility for nonspontaneous process to take place.

From the mathematical expression of thermodynamic coupling, i.e. [$\mathrm{d_i}S_1 < 0$, $\mathrm{d_i}S_2 > 0$ & $\mathrm{d_i}S \geqslant 0$], another natural rule, dissipation (or entropy production) decrease theorem, can be obtained. That is, in complex coupling systems, spontaneous irreversible process always inclines to utilize any useful energy in the form of transformation or storage into nonspontaneous process, i.e. negative entropy production process. The applicability framework of this theorem is just the same as that of modern thermodynamics, including nondissipative thermodynamics, linear and nonlinear dissipative thermodynamics for complex coupling systems. We can also say that this theorem is a generally applicable theorem, but in simple systems nonspontaneous process could not form because these systems are too simple. For instance, in a system composed of A and B gas mixture, heat conduction may cause a nonspontaneous diffusion against the concentration gradient; but in a single A gas system the heat conduction could not form a nonspontaneous diffusion. In fact, this theorem is the thermodynamic basis and guidance for biological evolution in nature. The energy transformation efficiency in living bodies is often high enough, and that is the result of the continuous complication and the continuous evolution. Therefore, this theorem is of a deep academic significance, and is a basic principle which should be followed in social science, progress in science and technology, and the development of society.

Dissipation decrease theorem was not found just now, because in Onsager’s famous article (1931) “Lord Rayleigh’s ‘principle of the least dissipation of energy’ ” had similarly been mentioned.[32, 33]

Both the extended Carnot theorem and dissipation (or entropy production) decrease theorem are the written statements (or connotations) of thermodynamic coupling, so they are also the written statements (or connotations) of the second law of modern thermodynamics. Classical thermodynamics emphasizes degradation of isolated systems, while modern thermodynamics emphasizes continuous evolution of nonequilibrium systems under a continuous energy-supply from the circumstance condition. Of course, modern thermodynamics for human beings living on the earth is of much deeper significance.

1.10 Some Fundamental Concepts and Definitions

Thermodynamics is a strict discipline of science. Any concepts introduced should be exactly defined. The definitions of some of the most fundamental terms in the thermodynamics discipline, such as "thermodynamics", "classical thermodynamics", "modern thermodynamics", and their applicability frameworks, should be very important. In this way, we know that thermodynamic coupling is the watershed between classical thermodynamics (or called traditional thermodynamics) and modern thermodynamics, and thermodynamic coupling is the core of modern thermodynamics.

1. Written Statement of the Thermodynamics Discipline. Thermodynamics is a part of science doing research on energy and energy transformation,[5] and its research objects are the change, development and evolution of macroscopic systems consisting of a large number of particles. The most important fundamental laws of thermodynamics are the first and the second laws of thermodynamics, i.e. "the first kind and the second kind of perpetual motion machines are impossible to obtain". In particular, the second law of thermodynamics is the core of thermodynamics. Up to now, "under any macroscopic conditions", including living body and biological evolution, the fact that the first kind and the second kind of perpetual motion machines have never been found demonstrates the generality and reliability of the thermodynamic theory, or we can say that the thermodynamic theory is really very rigorous.
2. Applicability Framework of the Thermodynamics Discipline. The applicability framework of the thermodynamics discipline is just the range of "under any macroscopic conditions". That is also the applicability framework of the first and the second laws of thermodynamics, or the applicability framework of "the first kind and the second kind of perpetual motion machines are impossible to obtain".
3. Written Statement of the Positive Entropy Production Principle. The entropy production of any systems is never negative.
4. Thermodynamic Coupling (i.e. Clausius' compensation). Thermodynamic coupling is a phenomenon that nonspontaneous process(es) may simultaneously take place together with spontaneous process(es) in complex multi-process systems. If chemical reaction(s) is (are) included in the system, then thermodynamic coupling is often called reaction coupling. The system including the phenomenon of thermodynamic coupling is called coupling system. The system without the phenomenon of thermodynamic coupling is called uncoupling system. Thermodynamic coupling is the watershed between clas-

sical thermodynamics and modern thermodynamics. Thermodynamics should not be divided into classical thermodynamics and modern thermodynamics merely based on time. Thermodynamic coupling is the core of modern thermodynamics.

5. Written Statement of the Second Law of Thermodynamics. The written statement of the first and the second law of thermodynamics may be "the first kind and the second kind of perpetual motion machines are impossible to obtain". Another more active statement can also be accepted, i.e. "Except the first kind and the second kind of perpetual motion machines, any kind of macroscopic energy transformation or any macroscopic state change of systems composed of a large number of particles is possible to realize." Thus, under activated conditions, low-pressure diamond growth and the appearance of living bodies under the circumstances similar to the earth biosphere are all thermodynamically possible. The first law of thermodynamics is also called the energy conservation law, so the first kind of perpetual motion machines does just violate the energy conservation law. The second kind of perpetual motion machines can be illuminated by practical examples of the change direction of a single-irreversible-process system. For instance, heat cannot be conducted from the colder body to the hotter body without any influence left, and so on. All of these statements on the change direction of a single-irreversible-process system can be regarded as equivalent. That is the common recognition of most thermodynamicists. However, all of these statements for single-irreversible-process systems are incomplete statements of the second law, and only suitable for simple uncoupling systems. Based on Clausius' "compensation", the written statement of the second law should be written in a complete and generalized form as follows. The second law of thermodynamics asserts that all spontaneous processes may take place (in a positive direction) by themselves without compensation (i.e. for simple uncoupling systems), such as heat conduction from the hotter body to the colder body; and any nonspontaneous process (in the negative direction) can only take place in such a manner as to be compensated by simultaneous occurring spontaneous process(es) (i.e. for complex coupling systems). Such a generalized statement could not be simply regarded as equivalent with a change direction of a single-irreversible-process system. The statement in the form of the second kind of perpetual motion machines is a short generalized form but without details.
6. The Mathematical Expression of Thermodynamic Coupling. Mathematical expression of thermodynamic coupling ≡ Mathematical expression of the second law of thermodynamics for complex (coupling) systems ≡ Mathematical expression of the second law of modern thermodynamics. All of their general mathematical expressions are the same, i.e. [$\mathrm{d_i}S_1 < 0$, $\mathrm{d_i}S_2 > 0$ & $\mathrm{d_i}S \geqslant 0$]. For isothermal isobaric systems, which are usually important in biochemistry, the mathematical expression of thermodynamic coupling could be written as

$$(\mathrm{d}G_1)_{T,p} > 0, (\mathrm{d}G_2)_{T,p} < 0 \ \& \ (\mathrm{d}G)_{T,p} \leqslant 0+ \tag{1.10}$$

because under isothermal isobaric conditions, Eq. (1.4) is $(\mathrm{d}G)_{T,p} = -T\mathrm{d_i}S$. The equality part of the mathematical expression of thermodynamic coupling corresponds to nondissipative thermodynamics (or nondissipative systems), while the inequality part of the mathematical expression of thermodynamic coupling corresponds to dissipative thermodynamics (or dissipative systems).

The detailed mathematical expressions of the second law for each field of thermodynamics have already been discussed in Section 1.4, and will not be repeated here.

1.11 Conclusion of this Outline

At the present, if "two physicists", as Truesdell said, or even if most scientists basically agree on this outline, that will be a tremendous success. Some differences may still be left in cognition. That will be a driving force for the further development of thermodynamics later. The development of the activated low-pressure diamond growth in the last 40 years has demonstrated the success of the thermodynamic coupling model and nondissipative thermodynamics of modern thermodynamics.

Thermodynamics is a macroscopic scientific theory, but not a microscopic one. Its research objects are macroscopic systems composed of a large amount (such as 10^{23}) of particles. The concepts or harvests of thermodynamics should not be extended to the whole of the universe or to the microscopic range with few particles.

In a single gas system, heat conduction cannot cause a nonspontaneous diffusion against concentration gradient and other "abnormal" phenomena. Among heat engines formed by simple uncoupling systems with cyclical processes, reversible heat engine is of the highest efficiency. That is the Carnot theorem. However, the necessary and sufficient condition for the highest conversion efficiency of energy is not "reversibility" but "nondissipation" of process or system. Therefore, nondissipative energy conversion is of the highest efficiency. That is the extended Carnot theorem.

The second law of thermodynamics is the core of thermodynamics. Some properties of thermodynamics and its applicability framework are just coming from the second law of thermodynamics, such as the irreversibility of the state change of macroscopic systems, i.e. the single-direction development characteristics in time of macroscopic systems. That is rarely found in the whole field of science, and thermodynamics is an important core part of science. The nature of the second law of thermodynamics is the "development" of macroscopic systems, and the "irreversibility" for the development of macroscopic systems, i.e. "arrow of time". However, human thinking was for a long time under the reign of classical thermodynamics. Some classical thermodynamicists had believed that it is not worthwhile to pay "more attention to irreversible phenemona, which are essentially transitory, than to the final result of their evolution, equilibrium."[36]

Some others also suggested that the English name of "thermodynamics" be changed into "thermostatics", because there was no time variable in classical (or traditional) thermodynamics. That is incorrect. Now there is also no time variable in the mathematical expression of the second law in modern thermodynamics, but the irreversibility of time for the development of systems is still incarnated.

As a science of development, the development of thermodynamics itself had been in stagnation for such a long time. That was rarely found in other fields of natural science. The thinking on development in thermodynamics had been "solidified" for such a long time. The mainstream on development should be a thinking of continuous "evolution", but it had been substituted by a thinking of "degradation". That was a pity, but is also an internal rule in the science of development.

The best answer to any difference in cognition may be to go forward based on the nature of the second law of thermodynamics, i.e. along the direction of “arrow of time”. History provides the best evidence. In the last 40 years of success in the activated low-pressure diamond growth, up to now as we know, only the thermodynamic coupling model and the corresponding nonequilibrium phase diagrams of modern thermodynamics have been evaluated as “well developed” and “of more strong persuasions” in thermodynamics.

References

1. Wang J-T. Progress in Natural Science [J], 2009, 19: 125-135. doi:10.1016/j.pnsc.2008.07.002.
2. Cápek V, Sheehan DP. Challenges to The Second Law of Thermodynamics: Theory and Experiment [M]. Berlin: Springer, 2004, (Truesdell’s words in p. 3).
3. Kleidon A, Lorenz RD (eds). Non-equilibrium Thermodynamics and the Production of Entropy – Life, Earth, and Beyond [M]. Berlin: Springer, 2005.
4. Lin Z-H (ed.). Thermodynamics and Statistic Physics [M]. Beijing: Beijing University Press, 2007, 1-2.
5. Holman J P. Thermodynamics [M]. 3rd edn. New York: McGraw-Hill, 1980, 1.
6. Wang J-T. Modern thermodynamics—A whole view of thermodynamics [M]. Shanghai: Fudan Univ Press, 2005.
7. Wang J-T, Zheng P-J. Extended Carnot Theorem. 19th Internat. Conf. on Chem. Thermodynamics [C]. Boulder, CO, USA: 30 Jul - 4 Aug 2006. http://www.symp16.nist.gov/pdf/p745.pdf. Cited 20 Apr 2008.
8. Clausius R. Mechanical Theory of Heat [M]. London: John van Voorst, 1867, Ninth memoir.
9. Wang J-T. Modern Thermodynamics in the Current 21st Century. 20th Internat Conf on Chem Thermodynamics [C]. Warsaw, Poland: 3-8 Aug 2008.
10. Wang J-T, Zheng P-J. The Second Law in Modern Thermodynamics. 19th Internat Conf on Chem Thermodynamics [C]. Boulder, CO, USA: 30 Jul - 4 Aug 2006. http://www.symp16.nist.gov/pdf/p2086.pdf. Cited 20 Apr 2008.
11. Wang J-T, Zheng P-J. Thermodynamics for Carat-size Synthetic Diamonds. 19th Internat Conf on Chem Thermodynamics [C]. Boulder, CO, USA: 30 Jul - 4 Aug 2006. http://www.symp16.nist.gov/pdf/p2082.pdf. Cited 20 Apr 2008.
12. Wang J-T. A New Field of Nonequilibrium Nondissipative Thermodynamics (plenary lecture). Abstracts of Russian International Conf on Chemical Thermodynamics (RCCT2005) [C]. Moscow, Russia: 31 Jun - 2 Jul, 2005.
13. Guggenheim E A. Modern Thermodynamics by the methods of Willard Gibbs [M]. London: Methuen, 1933.
14. Kondepudi D, Prigogine I. Modern Thermodynamics — From Heat Engines to Dissipative Structures [M]. New York: John Wiley & Sons, 1998.
15. Wang J-T, Zhang D W, Liu Z-J. Thermodynamic Coupling Model for Low Pressure Diamond Growth from the Vapor Phase [M]. Beijing: Science Press, 1998, reprinted in 2000.
16. Wang J-T. Phase Diagrams of Stationary Nonequilibrium States — Thermodynamics for Low Pressure Diamond Growth from the Vapor Phase [M]. Beijing: Science Press, 2000.

17. Wang J-T. Nonequilibrium Nondissipative Thermodynamics – With Application to Low-Pressure Diamond Synthesis [M]. Heidelberg: Springer, 2002.
18. Yarbrough W A, Missier R. Science [J], 1990, 247: 688-696.
19. Wang J-T, Carlsson J-O. Surf & Coat Tech [J], 1990, 43/44: 1-9; and presented at 8th Internat Conf on Thin Films [C]. San Diego, USA: 2–6 Apr 1990.
20. Wang J-T, Zheng P-J. Chin Sci Bull [J], Eng edn, 1995, 40(13): 1141-1143.
21. Gogotsi Y, Welz S, Ersoy D A, McNallan M L. Nature [J], 2001, 411: 283-286.
22. Wang J-T, Cao C-B, Zhang P-J. J Electrochem Soc [J], 1994, 141(1): 278-281.
23. Heimann B, Raiko V, Buck V. Refractory Metals & Hard Mater [J], 2001, 19(3): 169.
24. Pons M, Ballutaud D. Chemical Vapor Deposition [J], 2004, 10(3): 123.
25. Felske J D. Appl. Mechanics Review [J], 2004, 57(3): B16–B17.
26. Engineering Academy of China, Academy of China, Status Quo of Materials Development of China and Countermeasure for Going to the New Century, Consultation Program. Consultation Group on Nonnonferrous Metallic Materials (ed). Consultation Report on Nonnonferrous Metallic Materials [R]. Shanxi: Shanxi Science and Technology Publisher, 2001, 229.
27. Dai D-H, Zhou K-S (eds). Preparation Technology and Applications of Diamond Film Deposition [M]. Beijing: Metallurgy Industry Publisher, 2001, Chap 2 and 191.
28. Dai D-H, Zhou K-S, Yuan Z-H et al (eds). Surface Science and Technology of Modern Materials [M]. Beijing: Metallurgy Industry Publisher, 2004, 372-373.
29. Wang J-T. Modern Thermodynamics of the 21st Century and Nonequilibrium Phase Diagrams for Carat-size Synthetic Diamond. Proc of 13th National Symp on Phase Diagrams and China-Japan Joint Symp on Phase Diagrams [C]. Amoy, China: 9 Nov 2006, 245.
30. Wang J-T, The Second Law of Thermodynamics in the Current 21st Century (plenary lecture). 2007 National Symp on Thermodynamics and Statistics [C]. Yanji, China: 6 Aug 2007, 13.
31. Wang J-T. Modern Thermodynamics in the Current 21st Century. 5th Internat Sypm on Calorimetry and Thermal Analysis (CATS2008) [C]. Dailian, China: 18-21 May 2008.
32. Onsager L. Physical Review [J], 1931, 37: 405.
33. Onsager L. Physical Review [J], 1931, 38: 2265.
34. http://www.britannica.com/EBchecked/topic/215247/fourth-law-of-thermodynamics# [EB]. Cited 20 Sep 2008.
35. Prigogine I, Defay R. Chemical Thermodynamics [M]. translated by Everett DH. London: Longmans Green and Co Inc, 1954.
36. http://nobelprize.org/nobel_prizes/chemistry/laureates/1977/prigogine-autobio.html [EB]. Cited 20 Dec 2007.

Chapter 2

Brief Histories of Thermodynamics

Abstract An understanding of the development history of science is very important. In the field of thermodynamics the tendency of classicalism to refuse modernization is rarely found in other disciplines of science. No wonder that it was said at the end of the 20th century that [1]

> Thermodynamics is a core of science. Most books present thermodynamics very much the way it was presented in the 19th century.

In the development history of thermodynamics there are a lot of vivid examples and lessons which show this. The research objects have changed from the simplest reversible heat engines in Cartnot's time to complex systems including living systems in the present time, but the development history step by step in detail is still very useful for finding and removing out-of-date concepts and assumptions, which were merely suitable for simple systems.

2.1 Ancient Knowledge on Heat

The understanding on "heat" seems to start after one's birth. In fact, that is concerned with a history of several hundred years or longer. [1, 2]

Due to the sensory organs of human beings, we feel the difference between hot and cold. Studies on the nature of heat started in ancient times. In ancient Rome, there were several opinions on heat. In about 500 BC Heraclitus regarded fire as an element, which with water, earth, and gas were four independent elements in nature. The whole of nature might be composed of these four elements. In about 300 BC ZHOU Yan of Warring States in ancient China proposed a hypothesis of five elements ("wu xing"), that metal, wood, water, fire and earth composed the physical universe (everything in heaven and earth)[2]. Among them fire has a close connection with heat. In ancient China there was another hypothesis of male and female, that the physical universe were composed of two kinds of "qi". Fire was regarded as a performance of male "qi". They said: "Accumulation of male hot 'qi' becomes fire, and the essence of fire 'qi' is the Sun"[3]. Of course, these opinions had no scientific grounding enough to elucidate the nature of heat. In general, before the 18th century these heat elemental hypotheses, in which heat was regarded as an element for composing matter, were dominant. Based on some simple experiments

---

[1] http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471973947.html. Cited 2000–2002

[2] "Lu Shi Chun Qiu", 239 BC in China; "Hong Fan Pian" in "Shang Shu", before 239 BC in China.

[3] "Huai Nan Zi", 164 BC in China.

and some peripheral observations at this early time, the heat elemental hypotheses got its detailed statement: Heat was an indestructible fluid-like substance called caloric that was exchanged between material bodies with no procreation and no destruction. The hot or cold of an object was the performance of the containing caloric. Therefore, the heat elemental hypothesis might be called caloric theory of heat as well.

The opposite of caloric theory of heat was the mechanical theory of heat, or energy theory of heat. Friction producing heat was a well-known phenomenon in early times, so heat might be related with movement. Roger Bacon (1214–1294) believed that heat was a performance of motion. However, in general, only after the measurement of the equivalence between heat and work in the middle of the 19th century, could the correct viewpoint be established, that heat is a performance of energy. The caloric theory of heat entirely dropped out of the historical stage of science hereafter. Because the calorie or cal had already been adopted as the unit of heat, the unit of calorie was continuously used for a long time hereafter, until the unit of Joule was adopted as a new unit of heat (1 cal or calorie = 4.184 J or Joules).

2.2 Carnot Theorem

Carnot theorem: Between the same hot heat reservoir and the same cold heat reservoir the efficiency of reversible cyclical heat engines must be the maximum. A corollary of this conclusion may be: Between the same hot heat reservoir and the same cold heat reservoir the efficiency of all reversible cyclical heat engines must be the same regardless of their construction. Carnot theorem was proposed by Sadi Carnot (Fig. 2.1)[4] in 1824.

It is well-known that James Watt (Fig. 2.2)[5] innovated an applicable version of steam engines in the second half of the 18th century, and obtained a patent for the innovation of this kind of steam engines.[6] At that time he was working at the University of Glasgow as the most famous pupil and researcher under supervisor Joseph Black (1728–1799), who was a famous research scientist on heat and a professor of medicine and chemistry at the

Fig. 2.1 Sadi Carnot (1796–1832)

Fig. 2.2 James Watt (1736–1819)

---

[4] http://www-history.mcs.standrews.ac.uk/PictDisplay/Carnot_Sadi.html. Cited 17 Oct 2008

[5] http://www.egr.msu.edu/~lira/supp/steam/wattbio.html. Cited 17 Oct 2008

[6] http://www.infoscience.fr/histoire/biograph/biograph.php3?Ref=114. Cited 20 Apr 2008

university. This innovation of steam engines promoted an unimagined scale of industrial revolution, and a continuous pursuit to improve the efficiency of heat engines (stream engine is one kind of heat engines). Between May 1822 and September 1825, a railway from Stockton to Darlington in England (about 21 km length) was constructed with two trains using steam engine. That was the first railway in the world.[7] With such a background the appearance of Carnot theorem to get the highest efficiency of heat engines was natural and inevitable.

The appearance of Carnot theorem in 1920s was an important creative breakthrough in the development of modern science and technology. That made a foundation for a new discipline of thermodynamics, which was entirely different from Newton mechanics. If after 30 years the establishment of the first law and the second laws of thermodynamics was regarded as the sign of the birth of thermodynamics, then Carnot theorem was the start point of the thermodynamics discipline and especially the start point of the second law of thermodynamics. Sadi Carnot's father, Lazare Carnot (1753–1823), held many positions during and after the French Revolution and was known for his contributions to mechanics and mathematics. Lazare Carnot had a strong influence on his son Sadi. Sadi Carnot (1796–1832), a brilliant French military engineer, lived in rapidly industrializing Europe, he wrote in his memoir, (p. 68) [2], (p. 3) [3]

> Every one knows, that heat can produce motion. That it possesses vast motive-power no one doubt, in these days when the steam-engine is everywhere so well known.

It was his interest in general principles that led Sadi Carnot to his abstract analysis of steam engines.

Carnot considered the principles ruling the operation of steam engine. He first realized the flow of heat from high temperature to low temperature to be the driving force for producing "motive-power", i.e. "work" today. He pointed out: (p. 9) [3]

> Wherever there exists a difference of temperature, motive force can be produced.

That was similar to a water power machine, which works by a driving force coming from a water flow from high place to low place. Carnot understood that the restriction of doing work was independent of the machine and the manner of its operation, but only dependent on the temperature difference for producing the heat flow. He said: (p. 71) [2], (p.20) [3]

> The motion power of heat is independent of the agents employed to realize it; its quantity is fixed solely by the temperature of the bodies between which is effected, finally, the transfer of caloric.

He also believed that the work obtained from a certain amount of heat flow must be limited. He said: (p. 69) [2], (p. 13) [3]

> The necessary condition of the maximum (work) is, then, that in the bodies employed to realize the motion power of heat there should not occur any change of temperature which may not be due to a change of volume. Reciprocally, every time that this condition is fulfilled the maximum will be attained. This principle should never be lost sight of in the construction of heat-engines; it is its fundamental basis. If it cannot be strictly observed, it should at least be departed from as little as possible.

That is a reversible process with a minimum temperature gradient between the heat

---

[7] http://www.tworker.cn/bainian/firstway.htm. Cited 20 Apr 2008

reservoir and the heat engine.

Another important thought of Carnot was the cyclical operation of heat engines, so after doing work due to the heat flow the heat engine was recovered to its initial state, preparing to the next cycle. Carnot knew that the reversible cyclical heat engine must produce the maximum work. Otherwise, [3]

> This would be not only perpetual motion, but an unlimited creation of motion power without consumption either of caloric or of any other agent whatever. Such a creation is entirely contrary to ideas now accepted, to laws of mechanics and of sound physics. It is inadmissible.

The development of this theorem resulted in the birth of the second law of thermodynamics. The statement of Carnot theorem in the present textbooks may be as follows. Between the same hot heat reservoir and the same cold heat reservoir the efficiency of reversible cyclical heat engines must be the maximum. A corollary of this conclusion may be: Between the same hot heat reservoir and the same cold heat reservoir the efficiency of all reversible cyclical heat engines must be the same regardless of their construction.

Fig. 2.3 Emik Clapeyron (1799–1864)

In 1824, 28 years old Carnot described his general analysis of heat engines in his only scientific publication, "*Reflextions on the Motive Force of Fire and on the Machines Fitted to Develop that Power*". [3] Six hundred copies of this work were published at Carnot's own expense. At that time, the name Carnot was well known to the scientific community in France due to the influences of Sadi's father. However, Sadi Carnot's book did not attract much attention from the scientific community at that time. Eight years after the publication of this book, i.e. in 1832, 36 years old Sadi Carnot died of cholera. One year later, Emik Clapeyron (Fig. 2.3)[8] read Carnot's book and realized its fundamental importance. In 1834, Clapeyron wrote a paper "*An Account of Carnot's Theory of the Motive Power of Heat*" and published it in the *Journal de l'Ecole Polytechnique*. [4] In this paper Clapeyron reproduced Carnot's main ideas. Clapeyron represented Carnot's example of a reversible engine in term of a p (pressure, vertical coordinate) — V (volume, horizontal coordinate) diagram, as shown in Fig. 2.4. Clapeyron described the four single-process steps of Carnot cycle (i.e. adiabatic expansion, isothermal compression, adiabatic compression and isothermal expansion) with detailed mathematical treatments. Carnot cycle described in the manner of such a diagram must consist of reversible processes, and is used today in most textbooks of thermodynamics. [5]

Clapeyron's paper was later read by Lord Kelvin (1824–1907, and before 1892 his name was William Thomson) (Fig. 2.5)[9]. [6] In 1850, Rudolf Clausius (1822–1888) (Fig. 2.6)[10] on the basis of Clapeyron's and Kelvin's papers generalized the Carnot cycle into an arbitrary cyclical process, as shown in Fig. 2.7. (p. 89) [7] Before 1864, Clausius had no opportunity to read Carnot's original book. [8]

---

[8] http://en.wikipedia.org/wiki/%C3%89mile.Clapeyron. Cited 17 Oct 2008

[9] http://zapatopi.net/kelvin/quotes/. Cited 17 Oct 2008

[10] http://en.wikipedia.org/wiki/Rudolf_Clausius. Cited 17 Oct 2008

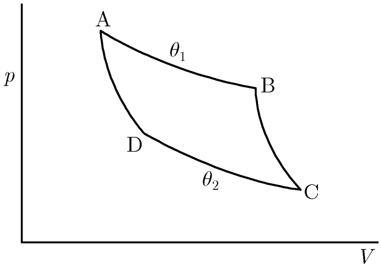


Fig. 2.4 The *p* — *V* diagram of Carnot cycle drawn by Clapeyron in his paper about Carnot's work

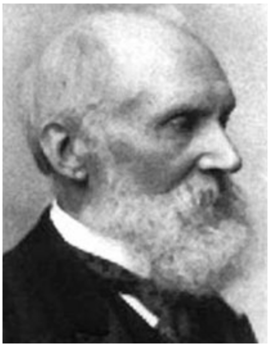


Fig. 2.5 Lord Kelvin (1824–1907)

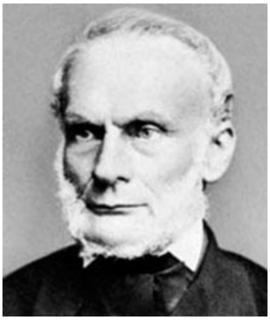


Fig. 2.6 Rudolf Clausius (1822–1888)

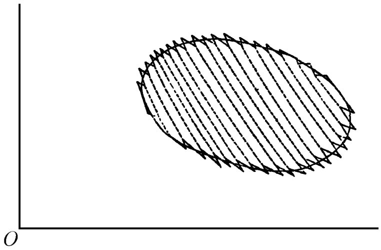


Fig. 2.7 Original figure of Clausius' generalization of Carnot cycle

Meanwhile, it should also be pointed out that the deduction or demonstration of Carnot theorem made by Carnot himself at that time was based on two conflicting theories (a correct theory of impossibility of the perpetual motion machines and an incorrect caloric theory of heat). The demonstration made by Carnot himself was as follows.

Supposing that heat engines A and B are working at the same hot heat reservoir to absorb heat Q_1 and Q_1', and to do work W and W' for their environment, respectively. Therefore, their efficiencies are η_A and η_B, respectively.

$$\eta_A = \frac{W}{Q_1} \text{ and } \eta_B = \frac{W'}{Q_1'}. \tag{2.1}$$

If A is a reversible heat engine, it is necessary to demonstrate $\eta_A \geqslant \eta_B$. For convenience, it is first supposed that $Q_1 = Q_1'$. If the theorem were not true, i.e. $\eta_B > \eta_A$, then one would get $W' > W$. Now it is known that heat engine A is a reversible engine, so the work W' done by the heat engine B might be utilized to force the reversible heat engine A to go backward. In this way the external work W were expended, and heat Q_1 might be laid back to the hot heat reservoir. The entire result of the combination of two heat engines A and B would be only an excessive work $W' - W$ left. Because all working matter of two heat engines had been recovered and the absorbed heat from the hot heat reservoir had already sent back with no change of the hot heat reservoir, an excessive work $W' - W$ was still left. The incorrect caloric theory of heat was used by Carnot at that time, so he considered that the heat absorbed from the hot heat reservoir was quantitatively equal to the heat ejected to the cold heat reservoir in amount, i.e. $Q_2 = Q_1$, $Q'_2 = Q'_1$. Therefore, Carnot considered that at the end of the working cycle of the combination of heat engines A and B, there were no change of cold heat reservoir left as well. Therefore, Carnot concluded that the superabundance of the work was obtained from nothing, so that would be a perpetual motion machine, which was known to be impossible, and a unique possible conclusion $\eta_\mathrm{A} \geqslant \eta_\mathrm{B}$ was obtained.

Now everybody knows that Carnot's conclusion, i.e. Carnot theorem, is correct, but the demonstration made by Carnot at that time was incorrect. Based on the first law of thermodynamics, after the cycle had been finished, all working matter must be recovered into the original state, so the internal energies must also come to their original values. Therefore,

$$W = Q_1 - Q_2 \text{ and } W' = Q'_1 - Q'_2. \tag{2.2}$$

Subtracting W' by W,

$$W' - W = (Q'_1 - Q_1) - (Q'_2 - Q_2). \tag{2.3}$$

It is known that $Q'_1 = Q_1$, so

$$W' - W = (Q_2 - Q'_2). \tag{2.4}$$

Therefore, the "superabundant" work $W' - W$ was coming from the decrease of heat $Q_2 - Q'_2$ in the cold heat reservoir, and was not coming from nothing as Carnot considered. It means that only based on the impossibility principle of energy coming from nothing, i.e. the impossibility principle of the first kind of perpetual motion machine or the first law of thermodynamics, which was established later, Carnot theorem cannot be demonstrated.

It was very interesting that a correct theorem had been "demonstrated" by an incorrect method. It was funny that "correctness might be established through incorrectness". From Carnot's later scientific notes it was known that Carnot himself later also gave up on the caloric theory of heat, because the caloric theory of heat was not supported by experiments. In fact, Carnot understood the mechanical equivalent of heat and even estimated the conversion factor to be approximately 3.7 joule per calorie later (the more accurate value being 4.18 J/cal). (p. 71) [2], [9] Unfortunately, Sadi Carnot's brother, Hippolyte Carnot, who was in possession of Sadi's scientific notes from the time of his death in 1832, did not make them known to the scientific community until 1878.

After giving up on the caloric theory of heat, it was found that there was no way to demonstrate the Carnot theorem on the basis of cognized principles at that time, and there was necessity to establish an entirely new law. Such a task was passed through Clapeyron's paper early or late to Kalvin and Clausius. Later for solving this task, Clausius and Kalvin proposed their own individual statements of the second law of thermodynamics. Carnot theorem was still the starting point for establishment of the second law of thermodynamics. On the other hand, once the second law of thermodynamics had been established, the second law of thermodynamics could be used for demonstration of Carnot theorem. It was only necessary to change the original assumption $Q_1 = Q_1'$ into $W = W'$, so the result $Q_1 - Q_1' = Q_2 - Q_2'$ would be obtained. If $\eta_B > \eta_A$, then $Q_1 > Q_1'$. At the end of the cyclical operation of the combined heat engines A and B, heat $Q_1 - Q_1'$ would be transferred from the cold heat reservoir to the hot heat reservoir. That would be a second kind of perpetual motion machine, so it was impossible. The unique conclusion left was $\eta_A \geqslant \eta_B$, as prescribed by Carnot theorem. The full import of the concepts originating in Carnot theorem was realized by Clausius through the introduction of the entropy function. Usually, the suggestions of the statements of the second law of thermodynamics made by Clausius and Kalvin, and especially the introduction of the entropy function S together with establishing the mathematical expression of the second law of thermodynamics, were regarded as the birth of a discipline of thermodynamics. Clausius and Kalvin were therefore regarded as founders of the thermodynamics discipline. However, some considered Carnot the real founder of thermodynamics. In fact, the appearance of Carnot theorem signaled the emergence of the thermodynamics discipline which then formed over several decades. So Carnot, Clausius and Kalvin are all founders of the thermodynamics discipline.

2.3 The Nature of Heat

Heat is a performance of microscopic motion and energy, and heat can be transformed to other forms of energy or vice versa. Temperature is a numerical value presentation of the hotness or coldness of a material object in its equilibrium state. Heat and temperature are two different basic conceptions. When two material objects are in heat equilibrium, i.e. there is no macroscopic change between two touched material objects for a long time, their temperatures must be equal. That is the zeroth law of thermodynamics, which we will talk about later, and the criterion or definition for temperature or temperature measurement. Thereby, the local equilibrium approximation was already introduced practically by Carnot, Clausius et al. during their discussions of material objects with different temperatures.

The development of modern science is based on experimentation. It is believed that nature obeys simple and universal laws, which we can understand and express in the precise formulations of mathematics. The key that unlocked this secret was experimentation and quantitative study of physical quantities (p. 7).[2] From the 16th century to the 17th century, Galileo Galilei (1564–1642) made an instrument on the basis of air expansion in heating, and it had been used in his scientific researches to show the temperature qualitatively. In 1660, Grossherzog Ferdinand II produced the first thermometer by putting alcohol in a glass tube. In 1714, Daniel Fahrenheit (1686–1736) established the Fahrenheit scale of temperature. He defined 0 °F as the temperature of the mixture of ammonium chloride and ice, and 100 °F as the temperature of human body. Since then the temperature was reported in the literature at different places and compared for academic exchanges and discussions. In 1742, Anders Celsius (1701–1744) set the freezing point of water as 100 °C and the boiling point of water as 0 °C, but this was at odds with common sense. His assistant Strömer reversed his scale of temperature, setting again the freezing point and boiling point of water as 0 °C and 100 °C, respectively. That is the most useful Celsius scale of temperature. In the 18th century, Joseph Black distinguished two fundamental concepts of heat and temperature by using the newly developed thermometer, and recognized a fundamental fact that, in thermal equilibrium, the temperature of all substances is the same. This idea was not easily accepted at that time because it seemed to contradict the ordinary experience of touch: a piece of metal feels colder than a piece of wood. However, the thermometer proved the point beyond doubt. Black also discovered the specific heat of substances and latent heats of fusion. The evaporation heat of water was obtained with the help of his pupil James Watt. (p. 9) [2, 10]

In the research mentioned above, the nature of heat remained a riddle for a long time. It was not known whether heat was an indestructible substance, called "caloric", which can move from one substance to another substance or whether heat was a form of microscopic motion and energy. Debates were continued to as late as the middle of the 19th century. Finally it became clear that heat was a form of energy that could be transformed to other forms or vice versa.

During 1744–1747, Mihail Lomonosov (Fig. 2.8)[11] believed that heat was a performance of molecular motion in his papers for discussion on the cause of hot and cold. Count Rumford (His original name was Benjamin Thompson, 1753–1814) is probably the first person who objected to the caloric theory of heat on the basis of experimental demonstration. In his paper published in 1798, he pointed out that in gun factories fragments of metals with very high temperature were continuously produced through the cutting process. It meant that the heat was continuously produced and the heat was converted from mechanical energy. He put metal cylinders in water and drilled holes in them. The heat produced due to mechanical friction made water boiled. Rumford estimated that for producing 1 calorie of heat about 5.5 joules are needed. He concluded that heat must be a performance of motion. (p. 33) [2] In the next year (1799), Humphry Davy (1778–1829) made more experiments to confirm the mechanical theory of heat. Ice could be melted only due to the friction between two pieces of ice. This experiment could not be explained by caloric theory of heat. These developments did not attract enough attentions of scientists at the time. In particular, the caloric theory of heat still prevailed in France. Antoine-Laurent Lavoisier (1743–1794), Juean-Baptiste-Joseph Fourier (1768–1830), Pierre-Simon de Laplace (1749–1827) and many other famous scientists believed the caloric theory of heat, i.e. heat was an indestructible fluid-like substance called caloric that was exchanged between material bodies with no procreation and no destruction. In Britain at the same time, Robert Boyle (1627–1691), Isaac Newton (1642–1684) and others believed that heat was the microscopic motion of particles, referred to as the mechanical theory of heat or the energy theory of heat. In 1847, James Prescott Joule (1818–1889) (Fig. 2.9)[12]reported his careful experimental measurements. His reproducible value of the

---

[11] http://www.euchems.org/distinguished/18thcentury/lomonosov.asp. Cited 17 Oct 2008

[12] http://en.wikipedia.org/wiki/James_Prescott_Joule. Cited 17 Oct 2008

mechanical equivalent of heat was 4.184 J/cal. In this way the first law of thermodynamics was thereby established. Correspondingly, the caloric theory of heat was a stage entirely removed from the development history of science.

Fig. 2.8 Mihail Lomonosov (1711–1765)

Fig. 2.9 James Prescott Joule (1818–1889)

Clausius at the beginning of his first memoir on thermodynamics sensitively pointed out Carnot's incorrectness on the nature of heat. (p. 14–89)[8] Clausius said: (p. 14–15)[8]

> Carnot proves that whenever work is produced by heat and a permanent alteration of the body in action does not at the same time take place, a certain quantity of heat passes from a warm body to a cold one; ⋯. The transmission Carnot regards as the change of heat corresponding to the work produced. He says expressly, that no heat is lost in the process, that the quantity remains unchanged; and he adds, "This is a fact which has never been disputed; it is first assumed without investigation, and then confirmed by various calorimetric experiments. To deny it, would be to reject the entire theory of heat, of which it forms the principal foundation."
>
> I am not, however, sure that the assertion, that in the production of work a loss of heat never occurs, is sufficiently established by experiment. Perhaps the contrary might be asserted with greater justice; that although no such loss may have been directly proved, still other facts render it exceedingly probable that a loss occurs. If we assume that heat, like matter, cannot be lessened in quantity, we must also assume that it cannot be increased; but it is almost impossible to explain the ascension of temperature brought about by friction otherwise than by assuming an actual increase of heat.

In general, heat is regarded as a disordered form of kinetic energy of a large number of particle motions. Molecules are in continuous motion, collide with each other and have random kinetic energy in directions and values. The average kinetic energy of molecules is $(mv_{\mathrm{avg}}^2/2) = 3kT/2$. Here, m is the mass of molecule, v_{avg} is the average speed of molecule, $k = 1.381\times10^{-23}$ J K^{-1} is the Boltzmann constant, and T is the absolute temperature. During the phase transformations, heat does not change the temperature of the body, but results in a transformation of the state. That was the classical picture of particle motion, and now we know that the fields and heat radiation should be considered as well. Moreover, based on the mass-energy transformation law proposed by Albert Einstein (1879–1955), i.e. $E = mc^2$, this should also be considered for the application of the conservation of energy in nuclear reactions. In this book, no further discussions about them will be made.

2.4 The First Law of Thermodynamics

The first law of thermodynamics, or the energy conservation law, is that it is impossible to obtain the first kind of perpetual motion machines. That is, the production of a machine to do work without any dissipation of external power and substance providing is impossible. For an isolated system, the internal energy of the system keeps constant, i.e. $(\Delta U)_{\text{iso}} = 0$ or $(\mathrm{d}U)_{\text{iso}} = 0$. That is the mathematical expression of the first law of thermodynamics for an isolated system.

Now we come back to talk about the 18th century. In the 18th century, the appearance of steam engines and the following industrial revolution in the entire of Europe led to a series of problems and requirements. It had been imagined that it was possible to produce a machine to do work without any dissipation of external power and substance providing, or a perpetual motion machine. Many investigators had declared their innovations, but none of these innovations could be confirmed by experimental check. In 1775, the Paris Academy of Science had to make an announcement rejecting any innovation application of perpetual motion machines. In fact, that was a primary form of the first law of thermodynamics. When Carnot demonstrated his theorem on the efficiency of heat engines, such an idea had already been used. As motioned above, at that time the imagined perpetual motion machine would be a machine to do work continuously without any necessity of external power and substance providing. Such a machine now could be called the first kind of perpetual motion machine. A brief statement of the first law of thermodynamics is that the production of the first kind of perpetual motion machines is impossible. The first law of thermodynamics is also called the energy conservation law, because the total energy of any system and its surroundings always keeps constant. Energy can transform from one form to another form, that is, a certain amount of one form of energy disappears together with the appearance of an equivalent amount of another form of energy. In other words, energy is impossible to create or to destroy in any isolated system.

The mechanical energy of a material body is the sum of the kinetic energy and the potential energy. The energy transformation is possible between the kinetic energy and the potential energy. If the factor of friction can be neglected, the sum of the kinetic energy and the potential energy will keep constant. These ideas had been set up as early as the time of Newton and Leibniz. The similarity made the first law of thermodynamics easier to accept. That provided a powerful support to the motion theory of heat. In 1847, the establishment of the mechanical equivalent of heat by Joule clearly confirmed that heat is a form of energy. The mechanical work could be converted to heat, and heat could also be converted to the mechanical work. After 1847, the concept of the conservation of energy had already been widely accepted. Around that time many scientists had made contributions from different viewpoints. For instance, as early as in 1840 Russian scientist Germain Henri Hess (1802–1850) (Fig. 10)[13] had already proposed the law of "constant summation of heats of reaction".[14] He pointed out: (p.36) [2]

> The amount of heat evolved during the formation of a given compound is constant, independent of whether the compound is formed directly or indirectly in one or in a series of steps.

---

[13] http://www.edu365.cat/aulanet/comsoc/Lab_quimica/quimics/Hess.htm. Cited 17 Oct 2008

[14] http://www.chemistry.co.nz/hess_law.htm. Cited 17 Oct 2008

That was an expression of the conservation of energy in chemical reactions. However, Russia was not at the center of the development of science and technology at that time. It was only after 40 years, when German scientist Wilhelm Ostwald (1853–1932) (Fig. 11)[15] made a citation of it in his book "*Textbook of General Chemistry*" published in 1887, that the law of "constant summation of heats of reaction" became known by chemists of other countries. This law is now well known as Hess law, which is very useful for calculating heats of chemical reactions. The conservation of energy as a general principle of nature was proposed first by a German doctor, Julius Robert von Mayer (1814–1878), in his paper "*Remarks on the Force of Inanimate Nature*" published in 1842, and a few years later by a German Physiologist, Hermann von Helmholtz (1821–1894), (Fig. 12)[16] in his paper "*On the Conservation on Force*" published in 1847. It was interesting that both Mayer and Helmholtz were not physicists at that time, so they were free from the caloric theory which prevailed in the physicist community. In the same year of 1847 the accurate value of the mechanical equivalent of heat was published by Joule.

Fig. 2.10 Germain Henri Hess (1802–1850)

Fig. 2.11 Wilhelm Ostwald (1853–1932)

Fig. 2.12 Hermann von Helmholtz (1821–1894)

The conservation of energy can be described and applied in terms of macroscopic variables of systems. A transformation of state may occur due to exchange of heat, performance of work and change in chemical composition. Each of them is associated with a change in energy. The first law of thermodynamics may be described as follows. When a transformation of state of a system takes place, the algebraic summation of the different energy changes (such as heats exchanged, work done, etc.) is independent of the path, i.e. the way of the transformation. It depends only on the initial and final states of the transformation. The macroscopic state is closely connected with the macroscopic variables of the system. For example, for a closed system with only absorption of heat and doing work to surroundings, the change of the internal energy of the system ΔU equals the difference of

---

[15] http://nobelprize.org/nobel_prizes/chemistry/laureates/1909/ostwald-bio.html. Cited 17 Oct 2008

[16] http://en.wikipedia.org/wiki/Hermann_von_Helmholtz. Cited 17 Oct 2008

them, i.e. $\Delta U = Q - W$. Here, ΔU is the difference between the internal energy of the final state U_2 and the internal energy of the initial state U_1, i.e. $\Delta U = U_2 - U_1$. In this way, a state function of internal energy U for characterization of the system was introduced through the first law of thermodynamics. It also means that thermodynamics is a discipline of science for studying the macroscopic changes of systems. From the statement that energy is impossible to create or to destroy in any isolated system, the mathematical expression of the first law of thermodynamics can be obtained in the form of $(\Delta U)_{\text{iso}} = 0$ or $(\mathrm{d}U)_{\text{iso}} = 0$. Here, the subscript of "iso" means the expression being suitable for isolated systems.

2.5 Absolute Scale of Temperature

The zero degree in the absolute scale of temperature is called the absolute zero degree. The existence of the absolute zero degree of temperature is an objective fact in nature. The importance of the absolute scale of temperature and the absolute zero degree could be found in the expression of efficiency of reversible heat engines in this Section and in the third law of thermodynamics later.

Clapeyron reproduced Carnot's main ideas on the efficiency of heat engines in a paper, which was published in the *Journal de l'Ecole Polytechnique* in 1834. Clapeyron's paper was later read by Lord Kelvin and others. On the basis of Carnot's work, it might be supposed that there are two reversible heat engines operating between the same hot heat reservoir and the same cold heat reservoir. The empirical temperatures of the hot heat reservoir and cold heat reservoir are θ_1 and θ_2, respectively. The efficiencies of these two heat engines are η_{A} and η_{B}, respectively. Based on Carnot theorem, the efficiency of these two reversible heat engines must be $\eta_{\mathrm{A}} = \eta_{\mathrm{B}} = \eta$, so the efficiency of heat engines η must only be a function of these two temperatures, that is

$$\eta = \frac{W}{Q_1} = 1 - \frac{Q_2}{Q_1} \tag{2.5}$$

and

$$\frac{Q_2}{Q_1} = F(\theta_1,\ \theta_2)\,. \tag{2.6}$$

Suppose that the third reversible heat engine is operating between empirical temperatures θ_3 and θ_1. This heat engine absorbs heat Q_3 at temperature θ_3, and releases heat Q_1 at temperature θ_1, so one obtains

$$\frac{Q_1}{Q_3} = F(\theta_3,\ \theta_1)\,. \tag{2.7}$$

Combining this heat engine with one of the previous reversible heat engines, heat Q_1 released at the temperature θ_1 by this heat engine is just balance out heat Q_1 absorbed at the temperature θ_1 by the previous heat engine, so a new reversible heat engine operating between empirical temperatures θ_3 and θ_2 is formed. The total result is

$$\frac{Q_2}{Q_3} = F(\theta_3,\ \theta_2)\,. \tag{2.8}$$

Eliminating Q_1, Q_2 and Q_3 from the above three Eq. (2.6), Eq. (2.7) and Eq. (2.8), the following equation is obtained.

$$F(\theta_1,\ \theta_2) = \frac{F(\theta_3,\ \theta_2)}{F(\theta_3,\ \theta_1)}. \tag{2.9}$$

θ_3 may be an arbitrarily appointed temperature, and appears only on the right side of the equation, so the terms related with θ_3 in numerator and denominator of the right side must be removable. Thus one can get

$$\frac{Q_2}{Q_1} = F(\theta_1,\ \theta_2) = \frac{f(\theta_2)}{f(\theta_1)}. \tag{2.10}$$

$f(\theta_2)$ and $f(\theta_1)$ are the same function of different empirical temperatures θ_2 and θ_1, respectively. The function is dependent on the selection of the scale of the empirical temperature, and independent of any material properties and the magnitude of heat Q. If now a new scale of temperature T, called absolute thermodynamic scale of temperature (or simply called, absolute scale of temperature), proportional to $f(\theta)$ is introduced, then one gets

$$\frac{Q_2}{Q_1} = \frac{T_2}{T_1}. \tag{2.11}$$

Finally, if the difference of the absolute temperatures between the boiling point and the freezing point of water are defined as 100, then the absolute scale of temperature is completely established. The introduction of the absolute scale of temperature was done by Kelvin. In memory of him the unit of absolute scale of temperature was nominated as Kelvin, and the sign of the unit is represented by K. Note that the introduced absolute temperature is always positive because the efficiency of heat engines is positive. Therefore, the efficiency of reversible heat engines and Carnot cycle may be represented as

$$\eta = \frac{W}{Q_1} = 1 - \frac{T_2}{T_1} = \frac{T_1 - T_2}{T_1}. \tag{2.12}$$

By the way, the term "thermodynamics" was also coined by Lord Kelvin in his 1849 publication "*An Account of Carnot's Theory of the Motive Power of Heat*", [6] and this term is used till now in the discipline. Thermodynamics (from the Greek "thermos" meaning heat and "dynamics" meaning power) [17] is a branch of natural science (concerned mainly with physics, chemistry and biology).

2.6 The Second Law of Thermodynamics

The general statement of the second law of thermodynamics is that the production of the second kind of perpetual motion machine is impossible. That is, state changes of macroscopic systems in nature are irreversible, i.e. the development along a single direction. For

---

[17] Oxford American Dictionary and http://www.informat.io/?title=thermodynamics. Cited 20 Apr 2008

instance, “heat cannot of itself, without the intervention of any external agency, pass from a colder to a hotter body”, or “it is impossible to obtain work by cooling a body below the lowest temperature of the system”, and so on. The nature of the second law of thermodynamics is “arrow of time”, and thermodynamics is a science concerned with development of the macroscopic world.

As mentioned above, the appearance of Carnot theorem met the necessity for the development of heat engines. The concept and design of Carnot cycle was very clever. The statement of Carnot theorem may be: Between the same hot heat reservoir and the same cold heat reservoir the efficiency of reversible cyclical heat engines must be the maximum. A corollary of the theorem may be: Between the same hot heat reservoir and the same cold heat reservoir the efficiency of all reversible cyclical heat engines must be the same regardless of their construction. Such a statement seemed to coincide very well with daily experiences and logical consequences of human being in the 19th century. However, the “demonstration” made by Carnot himself was on the basis of an incorrect caloric theory of heat at that time, and no one had found a correct way to demonstrate the Carnot theorem in about the following twenty five years. After the accurate value of the mechanical equivalent of heat had been published by Joule in 1847, the first law of thermodynamics, i.e. the energy conservation law, had already been generally accepted. However, the demonstration of Carnot theorem was still left unsolved. That led scientists to realize that an unknown and more fundamental law, i.e. the second law of thermodynamics, was just waiting to be established.

The second law of thermodynamics, similar to the first law of thermodynamics, was established on the basis of a large amount of human observation in relation to macroscopic phenomena and facts. The second law of thermodynamics and the first law of thermodynamics are derived from macroscopic experiences of human beings, and cannot be deduced from other more fundamental laws. The conclusions and deductions coming from thermodynamics are convincing because they agree with objective practice of human beings. It was powerfully proved that the second law of thermodynamics is really an important fundamental law of nature.

The second law of thermodynamics is a special law concerned with energy transformations, such as transformation from heat or internal energy to mechanical energy or other energy forms, and vice versa. Such a law summarizes the automatical selectivity in direction of processes concerned with phenomena of energy transformations. If along the automatically selected direction in a simple system with only one process the process may be going on without any other change left, then such a process is called spontaneous process. The process with an opposite direction of spontaneous process is called nonspontaneous process. For example, during the contact of two bodies at different temperatures, heat does always pass from a hotter to a colder automatically without other changes left. The opposite process cannot occur automatically in such a single-process system, i.e. heat cannot pass from a colder to a hotter automatically without other changes left in such a single-process system. Note that, a refrigerator may pass heat from a colder to a hotter, but it is not a spontaneous process, and the other change left is a dissipation of electric power. Moreover, the refrigerator is not a single-process system. For another example, friction producing heat may entirely transfer mechanical energy into heat and increase the homogeneous temperature of a body without other changes left. However, to transfer heat into mechanical energy with an automatic decrease of the homogeneous temperature of a body without other changes left is impossible. One of the basic phenomena in thermodynamics is the tendency toward equilibrium of any isolated system. Along such a characteristic direction the process in a single-process system is a spontaneous process. The second law of thermodynamics summarized all kinds of characteristic directions in single-process systems. For instance, gas may automatically expanse to vacuum without any other change left, and the opposite change will not take place automatically. Moreover, if solutions with different concentrations were put together, the solvent may automatically diffuse from the high concentration to the low concentration, until the solution becomes homogeneous. All these phenomena show the characteristic directions of spontaneous process in a single-process system.

A spontaneous process must be an irreversible process, but an irreversible process may not be a spontaneous process, because any practical processes, including nonspontaneous processes really taking place with energy providing in different ways, must be an irreversible process.

From these facts, we may understand that: (1) It is impossible for a spontaneous process to automatically revert to its initial state without any other change left. To judge the spontaneous characteristic direction of a process, it is only necessary to know the relationship between the final state and the initial state of the process. Therefore, it is possible to find a state function corresponding to the final state and the initial state. For instance, entropy, which was found by Clausius, is one of such state functions. (2) Different kinds of spontaneous processes of simple systems in nature are always connected with each other. From the characteristic direction of a spontaneous process the characteristic direction of another spontaneous process can be deduced. Therefore, various spontaneous processes of simple systems in nature can be described by individual examples. Clausius and Kelvin proposed their individual statements of the second thermodynamic law in 1850 and 1851, respectively. In fact, these statements were equivalent, i.e. they might be deduced from each other. Clausius statement of the second law of thermodynamics was that "heat cannot of itself, without the intervention of any external agency, pass from a colder to a hotter body." Kelvin statement of the second law of thermodynamics was that "it is impossible to obtain work by cooling a body below the lowest temperature of the system." The former described the characteristic direction of a spontaneous heat conduction, while the latter described the characteristic direction of a spontaneous heat production from friction. Once the Clausius statement was violated, i.e. if heat might spontaneously conduct from low temperature to high temperature, then heat would produce work during heat reflowing from high temperature to low temperature. Continuously repeating such a cyclical process for transformation from heat to work, heat would be taken from a single heat reservoir to produce work forever. That would be also a perpetual motion machine. Such a kind of perpetual motion machine is different from the first kind of perpetual motion machine, and does not violate the first law of thermodynamics, so it was called the second kind of perpetual motion machines. Such a written statement of the second law of thermodynamics that "it is impossible to produce the second kind of perpetual motion machine" may be more general. Sometimes this statement is called another Kelvin's statement.

The concept of "reversible process", as contrary to the basic concept of "irreversible process" had already been introduced in Carnot theorem. Reversible process is a process in which every step may be recovered backward without any other influences left. To describe the ideal and extreme meanings of reversible process, another concept of “quasistatic process” was introduced by Carathéodory.[11] The quasistatic process is an extreme slow process, which consists of a series of very close to each other equilibrium states. During the progress of each step of the quasistatic process the system is always in equilibrium. The reversible process is a quasistatic process without friction, and after the forward and backward without-friction quasistatic processes there is not any change of energy dissipation left, i.e. everything has been recovered. Therefore, “equilibrium state”, “nonequilibrium state” and “quasistatic process” are the most fundamental concepts of thermodynamics. In fact, both reversible process and quasistatic process are idealized processes, which may be approached but never be completely realized. Quasistatic process with friction is usually not considered in thermodynamics, so in thermodynamics sometimes quasistatic process, i.e. quasistatic process without friction, could be regarded as equivalent of reversible.

The nature of the second law of thermodynamics is the irreversibility of macroscopic systems in development. Such irreversibility need not be presented by a variable of time. The development irreversibility of macroscopic systems had already been described by Arthur Eddington using the words of “arrow of time”. This is important in order to catch hold of the nature of the second law of thermodynamics.

2.7 Entropy and Entropy Increase Principle

Entropy increase principle: Under adiabatic conditions, systems always tend to increase its entropy. Isolated systems are adiabatic systems, so one can say, entropy of an isolated system does never decrease, $(\Delta S)_{\text{iso}} \geqslant 0$ or $(\mathrm{d}S)_{\text{iso}} \geqslant 0$. Here, the subscript “iso” means for isolated (or adiabatic) systems. Both the introduction of entropy and the establishment of Clausius inequality represented the second law of thermodynamics coming into its quantitative stage in the form of mathematical expressions, and were also the sign of the formation of the thermodynamics discipline on the basis of the first and the second law of thermodynamics.

Once the second law of thermodynamics was established, there would be no difficulties for the demonstration of Carnot Theorem. Supposing that there were a heat engine with efficiency higher than a reversible heat engine, after combination of these two heat engines and running the reversible heat engine reversely, this combined set of heat engines would be a second kind of perpetual motion machine. The total result of this combined machine would be a continuous transformation of heat, coming from a heat reservoir with low temperature to produce work completely. That would be a violation of the second law of thermodynamics. However, the full import of the concepts originating in Carnot’s booklet was realized in the generalization made by Clausius and the introduction of a new concept of entropy, a physical quantity as fundamental and general as energy.

The first, from Eq. (2.11) $\frac{Q_2}{Q_1} = \frac{T_2}{T_1}$ one can obtain $\frac{Q_1}{T_1} - \frac{Q_2}{T_2} = 0$. For the purposes of mathematical treatment but not for discussion in written statements, it is not necessary to distinguish the heat absorption and the heat release. For the absorption of heat is Q, while for the release of heat is $-Q$. This rule will keep on in the following parts of the book.

In this way, the Q_2 in the equation above will be substituted by $-Q_2$. Therefore, from the Carnot cycle, the following equation may be obtained.

$$\frac{Q_1}{T_1}+\frac{Q_2}{T_2}=0. \tag{2.13}$$

Here, Q_1 and Q_2 are heats absorbed by the working matter at T_1 and T_2, respectively. According to the differential principle, any cyclical process may consist of a large amount of tiny Carnot cycles, as shown in Fig. 2.7. After addition or integration of all of equations of these tiny Carnot cycles, then one can obtain Eq. (2.14).

$$\sum\left(\frac{\delta Q}{T}\right)_{\mathrm{R}}=0 \quad \text{or} \quad \oint\left(\frac{\delta Q}{T}\right)_{\mathrm{R}}=0. \tag{2.14}$$

Here, δQ represents the heat absorbed by the working matter at temperature T, $\oint$ represents the rounding integration, and the subscript R represents the process being a reversible process. An arbitrary reversible cyclical process may be divided into two parts of reversible processes R_1 and R_2, and two connecting points may be called state A and state B, respectively, as shown in Fig. 2.13; then the following equation can be obtained.

$$\int_{\mathrm{A}}^{\mathrm{B}}\left(\frac{\delta Q}{T}\right)_{\mathrm{R}_1}+\int_{\mathrm{B}}^{\mathrm{A}}\left(\frac{\delta Q}{T}\right)_{\mathrm{R}_2}=0 \text{ or } \int_{\mathrm{A}}^{\mathrm{B}}\left(\frac{\delta Q}{T}\right)_{\mathrm{R}_1}=\int_{\mathrm{A}}^{\mathrm{B}}\left(\frac{\delta Q}{T}\right)_{\mathrm{R}_2}. \tag{2.15}$$

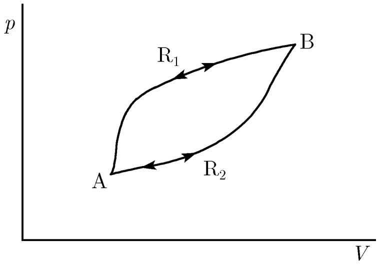


Fig. 2.13 Any reversible cyclical process may be divided into two parts of reversible processes.

Obviously, the sum of quotients of absorbed heat dividing by temperature is the same, and the value of the sum does not depend on any reversible path through which the change from state A to state B takes place. That is the characteristics of state function of a system. Therefore, Clausius defined a thermodynamics function of system, entropy, and denoted it by S. Its unit dimension is J K^{-1}. If entropy of the initial state and entropy of the final state of the system are represented by S_{A} and S_{B}, respectively, then one gets

$$S_{\mathrm{B}}-S_{\mathrm{A}}=\Delta S=\int_{\mathrm{A}}^{\mathrm{B}}\left(\frac{\delta Q}{T}\right)_{\mathrm{R}} \tag{2.16}$$

or

$$\Delta S-\sum\left(\frac{\delta Q}{T}\right)_{\mathrm{R}}=0. \tag{2.17}$$

If the equilibrium initial state and the equilibrium final states of the system are very close to each other, then the equation can be written in its differential form:

$$\mathrm{d}S = \left(\frac{\delta Q}{T}\right)_{\mathrm{R}} \quad \text{or} \quad \mathrm{d}S - \left(\frac{\delta Q}{T}\right)_{\mathrm{R}} = 0. \tag{2.18}$$

In 1865, Clausius in his famous paper which introduced this function said: (p. 357)[8]

> I propose to call the magnitude S the entropy of the body, from the Greek word $\tau\rho o\pi\eta$, transformation.

In the book "*Modern Thermodynamics*" written by Kondepudi and Prigogine and other books, we are often reminded that (p.80)[2]

> The usefulness of this definition depends on the assumption that any two states can be connected by a reversible transformation.

In fact, this is a severe restriction. In nature a large amount of macroscopic changes are always irreversible. The nature of the second law of thermodynamics is also the change irreversibility of macroscopic systems. That is to say, the framework of applicability of classical thermodynamics which was established on the bases of the definition of entropy and Carnot cyclical reversible processes, is very limited. Correspondingly, the history of the thermodynamics discipline from the proposal of Carnot theorem up to now about 180 years consists of the establishment and improvement of classical thermodynamics, which occurred mainly in the 19th century, and the exploration of modern thermodynamics for extending the applicability framework of thermodynamics, which occurred mainly in the 20th century and up to now.

According to Carnot theorem, the efficiency of any (including reversible and irreversible) heat engine working between the same hot heat reservoir and the same cold heat reservoir η' must be smaller or equal to the efficiency of reversible heat engine η, that is

$$\eta' = \frac{Q_1 + Q_2}{Q_1} = 1 + \frac{Q_2}{Q_1} \leqslant 1 - \frac{T_2}{T_1} = \frac{T_1 - T_2}{T_1} = \eta. \tag{2.19}$$

After transposition,

$$\frac{Q_1}{T_1} + \frac{Q_2}{T_2} \leqslant 0. \tag{2.20}$$

Extending this equation to so-called "arbitrary" process, in which there must be a reversible process connecting the equilibrium initial state and the equilibrium final state, the following equation can be obtained

$$\Delta S - \sum\left(\frac{\delta Q}{T}\right) \geqslant 0 \text{ or } \mathrm{d}S - \frac{\delta Q}{T} \geqslant 0. \tag{2.21}$$

This equation was usually called Clausius inequality, which was also a classical mathematical expression of the second law of thermodynamics.

For adiabatic processes, $\delta Q = 0$, so

$$\Delta S \geqslant 0 \text{ or } \mathrm{d}S \geqslant 0. \tag{2.22}$$

Eq. (2.22) was called the entropy increase principle, which was also a mathematical expression of the second law of classical thermodynamics. The statement of the entropy increase principle was: Under adiabatic conditions, the process for a system toward equilibrium always increases the entropy of the system. Any isolated system is adiabatic, so entropy of an isolated system does never decrease. That is another statement of the entropy increase principle. The corresponding mathematical expression should be

$$(\Delta S)_{\mathrm{iso}} \geqslant 0 \text{ or } (\mathrm{d}S)_{\mathrm{iso}} \geqslant 0. \tag{2.23}$$

The subscript "iso" means that the equation is suitable for isolated systems. Note that, scientific abstract and idealized extremum are very important in thermodynamics and other scientific disciplines. For instance, reversible process, quasistatic process and so on are typical examples. In an absolute sense, an isolated system does not exist. It is impossible that there is absolutely no influence and no connection between a system and its surroundings. The absolutely isolated system is impossible to be measured and to be recognized in physics. If the interaction between a system and its surroundings is much smaller than the internal change inside the system, then the system can be regarded as an isolated system.

2.8 Macroscopic Rules for Collective Motion of a Large Amount of Particles

The physical meaning of entropy is very difficult to understand in classical mechanics. Entropy is concerned with the collective behavior of a large amount of particles, so, entropy should be understood on the basis of statistical mechanics. $S = k \ln \Omega$ is the famous Boltzmann equation, which is written on his gravestone to this day as a result of his last wish before death. Here, k is Boltzmann constant, and Ω is thermodynamic probability, i.e. microscopic state number corresponding to macroscopic state. The Boltzmann equation also shows that entropy of thermodynamics is the statistical result of collective motions of a large amount of particles at the macroscopic scale. Therefore, this equation is fundamental to the neighboring discipline of statistical thermodynamics.

Applied steam engines were innovated in the second half of the 18th century and were used widely in industries at the beginning of the 19th century. This greatly promoted the birth and development of thermodynamics. At the time, some early developing modern sciences were also a necessary basis for the establishment of the thermodynamics discipline. In particular, the research objects of thermodynamics are the collective motion of a large amount of particles at the macroscopic scale (or called the change of system). In thermodynamics, the change of system can be deduced from macroscopic variables (such as temperature, pressure, volume, heat absorption and so on) of the system. The change of system is independent of the microscopic characteristics and structures of the internal particles, but it is necessary to know the initial state and the final state of the system. Such characteristics of thermodynamics are quite different from Newtonian mechanics, so it is impossible to research internal particles one by one using Newtonian mechanics. From Carnot theorem it is known that the efficiency of heat engine is independent of the working matter of the system. This led thermodynamics first to select the simplest group of a large amount of particles and the simplest reversible process as research objects. This is the important characteristic of classical thermodynamics. Among the three states of matter, i.e. solid, liquid and gas, gas is usually the simplest one. This was the reason for thermodynamics, during its initial establishment, to take ideal gas as the working matter of system and to discuss a single expansion or compression process.


Fig. 2.14 Robert Boyle (1627–1691)

Modern scientific research on gas may trace back to the 17th century. Robert Boyle (Fig. 2.14)[18] was living nearly at the same time as Isaac Newton (1642–1727). In 1660, in his paper, "*New Experiments Physico-Mechanical Touching the Spring of the Air and its Effects*", Boyle published his conclusion: At a constant temperature T, the volume of gas V was inversely proportional to the pressure p, i.e.

$$V = f(T)/p. \tag{2.24}$$

Here, f means a certain kind of function. This was usually called Boyle gas law. Of course, Boyle used the empirical temperature scale of his time. For convenience of discussion, the absolute temperature will be used in this book hereafter. Similar cases may be found in this book without any note. At nearly the same time Edme Mariotte (1620–1684) had also discovered this law in France.

About a century later, Jacques Charles (1746–1823) found the relationship between volume and temperature at constant pressure.

$$V/T = f'(p). \tag{2.25}$$

Here, f' is also a certain kind of function. In 1811, Amedeo Avogadro (1776–1856) (Fig. 2.15)[19] proposed his hypothesis: Under the same temperature and pressure conditions, equal volumes of gas contained an equal number of molecules. This hypothesis at that time could also be used for explanation of the pressure change in gas reactions. It meant under isothermal and isobaric conditions, the gas volume is proportional to the mole number N of the gas,

$$PV = Nf(T). \tag{2.26}$$

The above three equations could be summarized by the law of ideal gases, or the ideal gas equation.


Fig. 2.15 Amedeo Avogadro (1776–1856)

$$PV = NRT. \tag{2.27}$$

Here, R = 8.31441 J K^{-1} mol^{-1} is the gas constant.

---

[18] www.humanthermodynamics.com. Cited 20 Apr 2008

[19] http://www.bulldog.u-net.com/avogadro/avoga.html. Cited 17 Oct 2008

For a mixture of ideal gases there was also Dalton law of partial pressure, i.e. the partial pressure contributed by each gas component to the pressure of the mixture is only dependent on its own ideal gas equation, and independent of other gas components.

$$p_k V = N_k RT. \tag{2.28}$$

Here, p_k is the partial pressure of component k, and N_k is the mole number of gas component k. In 1807 French scientist Joseph-Louis Gay-Lussac (1778–1850) (Fig. 2.16)[20] discovered that temperature does not change during an expansion of a dilute gas into vacuum, and later in 1843 such a discovery was also confirmed by Joule.

However, it was also well-known that only the behavior of dilute gases obeys the law of ideal gases, i.e. Eq. (2.27). If the pressure was higher up to many atmospheres, the behavior of practical gases would deviate the law of ideal gases. In 1873 Dutch physics student Johannes van der Waals (1837–1923) (Fig. 2.17)[21] got his doctor's degree for his thesis entitled "*On the continuity of the gas and liquid state*", which made him immediately one of the famous physicists at that time. In this thesis he concluded that two main factors should be added to the ideal gas equation, i.e. the effect of molecular attraction and the effect of molecular size. The first factor makes the internal real pressure a little bigger than the apparent pressure p, and the second factor makes the internal real volume a little smaller than the apparent volume V. Therefore, he proposed an "Equation of State" for practical gases. That is,

$$(p + a/V^2)/(V - b) = N_k RT. \tag{2.29}$$

Fig. 2.16 Joseph-Louis Gay-Lussac (1778–1850)

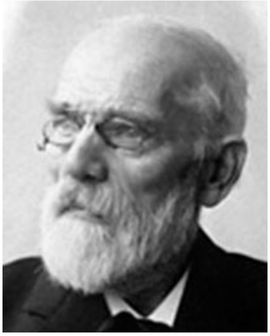


Fig. 2.17 Johannes van der Waals (1837–1923)

Here, two correction factors a and b had been introduced. In such a way his equation of state could be used for both the gaseous and the liquid states. The van der Waals equation could also provide a nice explanation for the critical temperature of real gases in gas-liquid phase transformation. Therefore, he demonstrated that these two aggregation states of matter can be continuously connected to each other, because the nature of them is of common point. The weak interaction force between molecules in gas, liquid and solid was called van der Waals force. Johannes van der Waals won the 1910 Nobel Prize in Physics "for his work on the equation of state for gases and liquids."

---

[20] http://www.answers.com/topic/joseph-louis-gay-lussac. Cited 18 Oct 2008

[21] http://nobelprize.org/nobel_prizes/physics/laureates/1910/. Cited 18 Oct 2008

The macroscopic rules of collective motion of a large amount of gas molecules provided thermodynamics with a lot of objective experiences and empirical bases. At the same time, the collective motion of gas molecules is just the research content of thermodynamics, so the research on the gas behavior was very important for the establishment of thermodynamics and its early development. For instance, the isothermal expansion of gas can be used for confirmation of the first thermodynamic law, $\Delta U = Q - W$; while the Gay-Lussac's gas expansion into vacuum clearly showed the irreversibility of collective motion of a large amount of particles. Gas may spontaneously expanse to vacuum, but gas can never automatically shrink form vacuum into a space part of the macroscopic container.

Thermodynamics studies the effects of changes in temperature, pressure and volume on system at the macroscopic scale, but these effects are due to the collective motion of a large amount of particles. The physical meanings of the function entropy of system (a collective behavior of a large amount of particles) is very difficult to be understood by any physical quantity in classical mechanics. Therefore, the function entropy should be understood on the basis of statistical mechanics. The Boltzmann equation $S = k\ln\Omega$ was a famous relationship proposed by Ludwig Boltzmann (1844–1906) (Fig. 2.18)[22] in 1896. Here, k is Boltzmann constant, and Ω is the thermodynamic probability with respect to the number of microscopic states. In this way, Boltzmann equation made a connection between thermodynamics and statistical mechanics, and that became the fundamental of statistical thermodynamics. For instance, put 4 balls (denoted as a, b, c, and d) into two boxes of the same volume (i.e. box 1 and box 2). The thermodynamic probability for finding the presentative state of 4 balls in one box is 2, because the 4 balls may be simultaneously found in box 1, and may also be simultaneously found in box 2. However, the thermodynamic probability for finding the presentative state of 2 balls in each box uniformly is 6 (i.e. 6 microscopic states). For a large number particles (such as, 6.023×10^{23} particles), these two thermodynamic probabilities for finding all particles in one box and finding particles uniformly in two boxes are quite different. The system always exists in the manner of the biggest thermodynamic probability. Therefore, gas molecules (particles) will quickly and uniformly be distributed in the whole container of box 1 and box 2, if a clapboard between box 1 and box 2 is removed. This is a spontaneous process of gas expansion to vacuum. It shows that a disordering process is always accompanied with an increase of entropy. This is also a relationship between disordering and entropy increase (or positive entropy production, which will be discussed later in this book).

Fig. 2.18 Ludwig Boltzmann (1844 – 1906)

Equilibrium statistics theory is established on the basic assumption of equal probability principle. Equal probability principle considers that for equilibrium state of isolated system the appearance probability of any possible microscopic state of the system is equal. Equal probability principle was also proposed by Boltzmann. Equilibrium statistics theory and its consequence, established on the basis of equal probability principle, had endured

---

[22] http://en.wikipedia.org/wiki/Ludwig_Boltzmann. Cited 18 Oct 2008

experimental verifications for more than one hundred years. It is not necessary to doubt its correctness, but it should not be aggrandized that the basic laws of thermodynamics were deduced from statistics. The basic laws of thermodynamics were established on the basis of a large number of human macroscopic experiences, but not deduced from any more basic law of other disciplines of science.

Note that, due to the success of thermodynamics some fundamental concepts, such as entropy, had also been extended into other fields. For example, the concept of entropy had been extended into the science of information where Shannon successfully developed the idea of entropy of information. Therefore, some attempted to list 21 kinds of "entropy" and 21 kinds of "the second law of thermodynamics", and wrote a book of "*Challenges to the Second Law of Thermodynamics*" in 2004.[12] In fact, Boltzmann equation and Shannon's entropy are neither the second law of thermodynamics itself, nor the contents of thermodynamics. The mistakes made by authors of that book might only be for the purpose of proposing their own opinions, which might belong to other kinds of disciplines. The so-called "quantum thermodynamics", "black-hole thermodynamics", "nanothermodynamics" and so on might be already out of thermodynamics and not belonging to the discipline of thermodynamics.

2.9 Development and Limitation of Classical Thermodynamics

The whole of classical thermodynamics was set up on the basis of Carnot theorem, i.e. classical thermodynamics considered practically only the simplest spontaneous-process systems and equilibrium systems without any macroscopic process. In following chapters, it will be known that classical thermodynamics only considered uncoupling systems, i.e. equilibrium systems and systems including only spontaneous process(es).

As discussed in the last section, thermodynamics studies the effects of changes in temperature, pressure and volume on systems at the macroscopic scale, and from a microscopic viewpoint the effects of changes are due to the collective motion of a large amount of particles. The microscopic viewpoint is very useful for deeply understanding the nature of heat and the characteristics of thermodynamics. In fact, discussions on ideal gas and rules of the changes in single processes of ideal gases are all based on macroscopic experiments and macroscopic measurable state variables, such as temperature, pressure, volume, moles of the quantity of matters, and so on. The first law and the second law of thermodynamics summarized a large amount of macroscopic experiences coming from our daily life, scientific experiments and productions of matter. After the introduction of some necessary assumptions and premises, the first law and the second law of thermodynamics had been expressed in classical quantitative mathematical expressions. Through these expressions, particularly on the basis of the equality of the second law together with the equality of the first law, a rather rigorous discipline of classical equilibrium thermodynamics had been deduced. That may be an outstanding example in history of science in its development.

In regard to the deduction of classical equilibrium thermodynamics, the contribution made by American scientist Josiah Willard Gibbs (1839–1903) (Fig. 2.19)[23] was extraor-

---

[23] http://en.wikipedia.org/wiki/Willard_Gibbs. Cited 18 Oct 2008

dinary.[13] Until today much of the content of textbooks on classical thermodynamics trace back to his works. Gibbs has been regarded as one of the greatest scientists of the 19th century. He was one of the few American physicists in the 19th century to get an international reputation and he was the only person to make a theoretical contribution of fundamental importance. In 1875 and 1878 Gibbs published his two scientific papers "*On the Equilibrium of Heterogeneous Substances*", which was his most important contribution to physical sciences. for an exhaustive study of geometrical methods of representing by diagram the thermodynamic properties of homogeneous substances. These early papers had attracted the attention of England's leading physicist, James Clerk Maxwell (1831–1879) (Fig. 2.20)[24] who constructed a model illustrating a portion of Gibbs work and sent a plaster cast to Gibbs. After then Gibbs published his third paper. It was a pity that Maxwell died soon after the publication of Gibbs' third paper, so Gibbs was still unknown in Europe. These important papers by Gibbs were published in the *Transactions of the Connecticut Academy of Sciences*, a journal edited by Gibbs' librarian brother-in-law, but the journal was little read in the USA and much less in Europe. "Gibbs Phase Rule" is now well known to all in thermodynamics and even in physics and chemistry. Nevertheless it was a number of years before its value was generally known. Such a delay was mainly due to causes from two sides. First, its mathematical form and rigorous deductive processes were difficult to understand for readers at that time. Second, in the 1870s, America was not the center of science and technology. Therefore, only after Gibbs' papers were translated into German (the leading language for chemistry at that time) by Wilhelm Ostwald in 1892 and into French by Henri Louis le Chatelier in 1899, were Gibbs' ideas widely accepted in Europe. Gibbs died few years later, and he did not win the Nobel Prize. To monumentalize the contributions of Gibbs, the state function of isothermal isobaric free energy is called Gibbs function G.

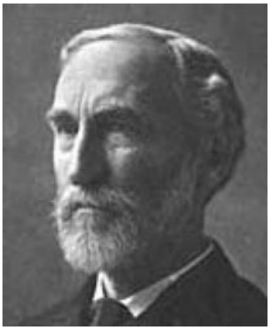


Fig. 2.19 Josiah Willard Gibbs (1839–1903)

Fig. 2.20 James Clerk Maxwell (1831–1879)

According to the expression of the second thermodynamic law, the change of entropy of a system may be calculated, but the calculation of the absolute value of entropy should recur to the third law of thermodynamics. In 1906, Walther Nernst (1864–1941) (Fig. 2.21)[25] proposed the Nernst heat theorem that at the absolute zero of temperature

---

[24] http://en.wikipedia.org/wiki/James_Clerk_Maxwell. Cited 18 Oct 2008

[25] http://nobelprize.org/nobel_prizes/chemistry/laureates/1920/nernst-bio.html. Cited 18 Oct 2008

the entropy of every chemically homogeneous solid or liquid body has a zero value. (p. 85) [14] The heat theorem was also called the third law of thermodynamics. Walther Nernst won the 1920 Nobel Prize in Chemistry "in recognition of his work in thermochemistry". The third law of thermodynamics is also called the principle of impossible to reach the absolute zero, i.e. one can never lower temperature of a body to the absolute zero by limited steps. From the efficiency expression of Carnot reversible heat engines, i.e. Eq. (2.12), it is known that if there were a heat reservoir at absolute zero, then one could take heat from any single heat source (such as the ocean) to completely transform it into work. Therefore, the third law of thermodynamics is also related with the second law of thermodynamics. This might be the reason that many thermodynamics textbooks, and this book, discuss mainly the first and the second laws of thermodynamics.

Fig. 2.21 Walther Nernst (1864–1941)

From the 19th century till now, the research and application of equilibrium thermodynamics has still been developing, such as the computer assistant calculation of phase diagrams (CALPHAD) which has played an important role in research into equilibrium phase diagrams. However, at the end of the 19th century, classical thermodynamics, through rigorous deduction made by Gibbs, James Clerk Maxwell and others on the basis of the first and the second laws of thermodynamics (especially on the basis of their equality parts) came into its mature stage. So in the early to mid 20th century the term classical equilibrium thermodynamics gradually appeared.

It was worth mentioning that Ostwald discovered Hess' and Gibbs' research contributions one by one in Russian and in America, respectively. At the same time, due to his own contribution Wilhelm Ostwald won the 1909 Nobel Prize in Chemistry "in recognition of his work on catalysis and for his investigations into the fundamental principles governing chemical equilibria and rates of reaction".

The confusion between "classical thermodynamics" and "equilibrium thermodynamics" can still be found today. That is not good because there are two parts of equality and inequality in the second law of thermodynamics, since the Clausius inequality and the entropy increase principle were proposed by Clausius. If the equality part may correspond to equilibrium thermodynamics, then the inequality part may correspond to nonequilibrium thermodynamics. The nonequilibrium thermodynamics (i.e. the inequality part of the second law) in classical thermodynamics could only be a criterion for some changes including irreversible processes. The inequality part of the second law in classical thermodynamics did not develop well. This agrees with the general rule of the cognition development for human beings: from simple to complex, from equilibrium to nonequilibrium, but it also due to the influence of out-of-date thought, as was said by Prigogine: [26]

> I remember that towards the end of 1946, at the Brussels IUPAP meeting, after a presentation of the thermodynamics of irreversible processes, a specialist of great repute said to me, in substance: "I am

---

[26] http://nobelprize.org/nobel_prizes/chemistry/laureates/1977/prigogine-autobio.html. Cited 18 Oct 2008

> surprised that you give more attention to irreversible phenomena, which are essentially transitory, than to the final result of their evolution, equilibrium."

The proposition of "the limitation of classical thermodynamics" or "the applicability framework of classical thermodynamics" did not attract the attention of most scientists in the thermodynamics forum. However, such a proposition was famously suggested by Albert Einstein in 1949: [15]

> A theory is more impressive the greater the simplicity of its premises is, the more different kinds of things it relates, and the more extended its area of applicability. Therefore the deep impression which classical thermodynamics made upon me. It is the only physical theory of universal content concerning which I am convinced that, within the framework of the applicability of its basic concepts, it will never be overthrown.

There was no detailed answer on what is "the limitation of classical thermodynamics" or "the applicability framework of classical thermodynamics" both at that time and for a long time after.

The limitation of classical thermodynamics has existed since the establishment of classical thermodynamics. Just as shown in Fig. 2.22, the research object of Carnot cycle was the simplest gas system, and even the ideal gas system. Therefore, the core of Carnot theorem was "reversibility" and "cyclical process", but the nature of the second law of thermodynamics is the irreversibility of macroscopic processes, described as "arrow of time" by Arthur Eddington. As said before, Clausius introduced entropy and established the mathematical expression which fully opened out the profound meanings of Carnot theorem. Therefore, the research objects in classical thermodynamics are simple systems, or more suitably to say, uncoupling systems.

Clausius in his famous 1865 paper, "*On several convenient forms of the fundamental equations of the mechanical theory of heat*", had to definitely introduce a cyclical-process assumption for getting the classical mathematical expression of the second law of thermodynamics "in the simplest manner". "If different parts of the body have different temperature," the local equilibrium approximation was also introduced. He wrote: (p. 328–329) [8]

> In order to express analytically the second fundamental theorem in the simplest manner, let us assume that the changes which the body suffers constitute a cyclical process, whereby the body returns finally to its initial condition. By $d\mathrm{Q}$ we will again understand an element of heat absorbed, and T shall denote the temperature, counted from the absolute zero, which the body has at the moment of absorption, or, if different parts of the body have different temperature, the temperature of the part which absorbs the heat element $d\mathrm{Q}$. If we divide the thermal element by the corresponding absolute temperature and integrate the resulting differential expression over the whole cyclical process, then for the integral so formed the relation
>
> $$\int \frac{d\mathrm{Q}}{\mathrm{T}} \leqslant 0$$
>
> holds, in which the sign of equality is to be used in cases where all changes of which the cyclical process consists are reversible, whilst the sign $<$ applies to cases where the changes occur in a nonreversible manner. *

(* $d\mathrm{Q}$ is positive for heat absorbed by a changing body and is negative for heat given off by it.) [Note: emphases were made by Clausius himself.]

Besides, during the deduction of Clausius inequality, as shown in Eq. (2.21), Clausius said:

> Generalizing for an arbitrary process, in which its initial and final states might also be connected with a reversible process, then the following equation was obtained.

Of course, the demand of "in which its initial and final states might also be connected with a reversible process" is also a severe restriction to the so-called "arbitrary process", because system in a reversible process is always in its equilibrium states. Hereafter, this restriction will be called the assumption of initial and final equilibrium states for the Clausius' "arbitrary process". For quantitative calculation of the entropy change from a standard state 0 to a state X, Eq. (2.16) may be rewritten as follows.

$$S_X = S_0 + \int_0^X \left(\frac{\delta Q}{T}\right)_R . \tag{2.30}$$

The system changes through an irreversible process from the standard state 0 to the state X, but the calculation must be on the basis of a reversible process R. In the book "*Modern Thermodynamics*" written by Kondepudi and Prigogine it was clearly pointed out that

> In classical thermodynamics it is assumed that every irreversible transformation that occurs in nature can also be achieved through a reversible process for which [Eq. (2.30), here] is valid.

Figure 2.22 illustrates the cyclical-process assumption and the assumption of initial and final equilibrium states for irreversible process in classical thermodynamics.

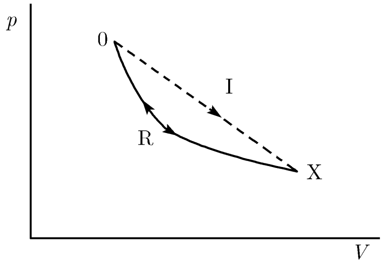


Fig. 2.22 The cyclical-process assumption and the assumption of initial and final equilibrium states for irreversible process in classical thermodynamics.

In Fig. 2.22 it can also be found that the cyclical-process assumption is closely connected with the assumption of initial and final equilibrium states. In Fig. 2.22, the process "I" together with the up arrow of the reversible process "R" forms a cyclical process, while the process "I" together with the down arrow of the reversible process "R" illustrates that the Clausius' "arbitrary process" under consideration in classical thermodynamics must be of initial and final equilibrium states. That is, in classical thermodynamics the entropy change due to the so-called "arbitrary process" must be equal to entropy change of the system via the reversible process. Such a severe restriction in classical thermodynamics will be further discussed in the next chapter.

In nature there are a lot of irreversible processes which are not cyclical processes. For instance, it is impossible to return the earth's biosphere together with the sun back to any pre-existing state. Therefore, it is impossible to deal with a large amount of irreversible processes in nature through classical thermodynamics. That is the same for the phenomena of life, which is always connected with nonequilibrium states. There is no life in equilibrium states, so the phenomena of life are generally impossible to be dealt with by classical thermodynamics.

In Clausius' original paper, the calculation of entropy change of "arbitrary process" through reversible process was expressed by "if" clause. For instance, Clausius said: (p. 355)[8]

> Denoting this magnitude by S, we can write
>
> $$dS = \frac{dQ}{T};$$
>
> or, if we conceive this equation to be integrated for any reversible process whereby the body can pass from the selected initial condition to its present one, and denote at the same time by S_0 the value which the magnitude S has in that initial condition,
>
> $$S = S_0 + \int \frac{dQ}{T}.$$

Such an "if" clause practically introduced the assumption of initial and final equilibrium states. Although this is a strongly restrictive condition it still may be realized in some cases.

However, at the end of Clausius' paper, he used another "if" clause to extend the first and the second fundamental laws of thermodynamics into so-called "the fundamental laws of the universe". He said: (p. 365)[8]

> If for the entire universe we conceive the same magnitude to be determined, consistently and with due regard to all circumstance, which for a single body I have called entropy, and if at the same time we introduce the other and simpler conception of energy, we may express in the following manner the fundamental laws of the universe which correspond to the two fundamental theorems of the mechanical theory of heat.
> (1) The energy of the universe is constant.
> (2) The entropy of the universe tends to a maximum.

This "if" clause at the end of Clausius' paper could not match his previous cyclical-process assumption and the assumption of initial and final equilibrium states, because it is impossible to demonstrate the change of the universe being a "cyclical process", and it is also impossible to demonstrate that the universe were of an initial equilibrium state and a final equilibrium state as well. Therefore, we should not use the incorrect term "the fundamental laws of the universe" or the incorrect term "entropy of the universe" in textbooks on thermodynamics anymore.

2.10 Exploration of Modern Thermodynamics in the 20th Century

The general mathematical expression of the second law of thermodynamics is the positive entropy production principle, $d_iS \geqslant 0$, which appeared around the beginning of the 20th century. The increase of entropy of the system dS has been divided into two parts of entropy production d_iS and entropy flow d_eS, i.e. $dS = d_iS + d_eS$. Here, entropy production is the part of entropy increase due to internal irreversible process(es) of the system, while entropy flow is the part of entropy increase due to exchanges of matter and energy between the system and its surroundings. Onsager and Prigogine won their Nobel Prizes in chemistry in 1968 and 1977, respectively. They made contributions to the development of researches on complex systems. In 1931, an out-of-thermodynamic assumption was introduced, and approximative On sager reciprocal relations were obtained. In the 1950s and 1960s the theory of dissipative structures was proposed on the basis of kinetic (out-of-thermodynamics) models by Prigogine. Just in 1931, a biochemist D. Burk proposed an important concept of thermodynamic coupling or coupled reactions, which would develop to be the core of modern thermodynamics, directly based on the second law of thermodynamics. In 1961 another biochemist P. Mitchell proposed a chemiosmotic coupling hypothesis for ATP biosynthesis, and won the 1978 Nobel Prize in chemistry. All of their contributions are important for developing modern thermodynamics.

The deduction in thermodynamics is usually very rigorous. The questions, mistakes or oversights might be introduced unconsciously. For instance, the statement of entropy increase principle in isolated system may be that "Entropy of an isolated system will never decrease" and the corresponding mathematical expression is $(\mathrm{d}S)_{\mathrm{iso}} \geqslant 0$. That seems absolutely no problem, but under careful consideration some additional explanations should be made. If the initial state of the isolated system is in equilibrium, there must be no internal macroscopic changes, i.e. $(\mathrm{d}S)_{\mathrm{iso}} = 0$, but if the initial state of an isolated system is not in equilibrium, then there must be no meaning of entropy value for the isolated system, because in classical thermodynamics the definition of entropy is only for system in equilibrium state. In practice, if the internal temperature gradient of the nonequilibrium state is not too high, then $(\mathrm{d}S)_{\mathrm{iso}} > 0$ could still be regarded as the correct one. This is concerned with the zeroth law of thermodynamics and the local equilibrium approximation. Some special cases, in which the changes from an equilibrium initial state to an equilibrium final state of isolated systems are switched by a valve or by drawing a clapboard away, will be discussed in the next chapter.

If system A is in equilibrium with system B and if system B is in equilibrium with system C, then one can say that system A is in equilibrium with system C. That may be called the zeroth law of thermodynamics. The zeroth law of thermodynamics requires thermometer being in equilibrium with the system for measurement, so there are no meanings of temperature and other thermodynamic parameters in their original rigorous definitions for nonequilibrium systems. The simplest solution to such a problem is the local equilibrium approximation. According to the local equilibrium approximation the nonequilibrium system is divided into a large number of small volumes, and each volume is small enough in macroscopic point of view, but still includes a large amount of particles in microscopic point of view. Once the small volume is isolated it will soon be in equilibrium much faster than the change of the system, so thermodynamic parameters inside the small volume after equilibrium could be regarded as thermodynamic parameters of the small volume. In this way, thermodynamic researches on systems without high temperature (or other thermodynamic parameter) gradient could be made, getting very nice results. Sometimes systems with a little higher temperature (or other thermodynamic parameter) gradient could still be treated similarly, but in such cases the results might have more qualitative meanings.

The application mathematical expression of entropy increase principle $(\mathrm{d}S)_{\mathrm{iso}} \geqslant 0$ should be limited only for isolated (or adiabatic) systems. Even after the introduction of state functions of Gibbs free energy G, Helmholtz free energy and the introduction of chemical potential made by Gibbs, Helmholtz and G.N. Lewis (1875–1946) (Fig. 2.23)[27], the application of classical thermodynamics was still mainly limited to isolated and closed systems. To find the more general mathematical expression of the second law of thermodynamics Pierre Duhem (1861–1916) (Fig. 2.24)[28], L. Natanson and G. Jaumann introduced concepts of entropy production and entropy flow. (p. 87) [2] The entropy increase of system $\mathrm{d}S$ is divided into two parts of entropy production $\mathrm{d_i}S$ and entropy flow $\mathrm{d_e}S$, i.e. $\mathrm{d}S = \mathrm{d_i}S + \mathrm{d_e}S$. Here entropy production is the part of entropy increase due to internal irreversible process(es) of the system, while entropy flow is the part of entropy increase due to exchanges of matter and energy between the system and its surroundings. A more general positive entropy production principle (which can be used for any macroscopic isolated, closed and open system) than the entropy increase principle was then obtained. The statement of the positive entropy production principle is: Entropy production of any system is never negative. The general mathematical expression of the second law of thermodynamics can be written as

$$\mathrm{d_i}S \geqslant 0. \tag{2.31}$$

The positive entropy production principle made a further generalization of the mathematical expression of the second law of thermodynamics and the local equilibrium approximation was included as well. The importance of the concept of entropy production is also due to its direct connection with the internal irreversible process(es).

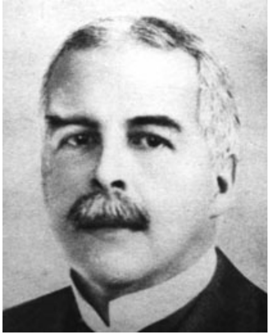


Fig. 2.23 Gilbert Newton Lewis (1875–1946)

Fig. 2.24 Pierre Duhem (1861–1916)

The term “classical thermodynamics” had probably appeared in the literature sometime in the first 3 decades of the 20th century. The original meaning of “classical thermodynamics” is likely to include the function of “entropy”, strictly on the basis of Clausius definition, and a series of rigorous deductions mainly in the simplified equality part made by Gibbs and others in the 19th century. All of these researches had already composed a rather complete discipline of science. Of course, the applicability of classical thermodynamics should be restricted. The mathematical deduction and logic consequence in classical thermodynamics were on the basis of the classical mathematical expressions of the first and the second laws of thermodynamics. In early development the term “thermodynamics” was sufficient. It was not necessary to call it “classical thermodynamics” at that time.

---

[27] http://www.eoearth.org/article/Lewis,_Gilbert_Newton. Cited 18 Oct 2008

[28] http://en.wikipedia.org/wiki/Pierre_Duhem. Cited 18 Oct 2008

With the development of science, at the beginning of the 20th century a lot of new phenomena were observed which could not be explained by thermodynamics (i.e. classical thermodynamics), so some approximation and extension or development had to be introduced. To distinguish the new exploration in thermodynamics, the term "classical thermodynamics" emerged. The appearance of the term "classical thermodynamics" signalled that non-classical thermodynamics or modern thermodynamics was beginning to be explored at this time in the early 20th century or later.

Fig. 2.25 Lars Onsager (1903–1976)

In 1931 Brown University's Lars Onsager (1903–1976) (Fig. 2.25)[29] proposed the first general approximatively quantitative equation — Onsager reciprocal relations in nonequilibrium thermodynamics. This was based on the "assumption of microscopic reversibility" or the assumption of "detailed balancing". [16] Onsager reciprocal relations elucidated that in complex systems, including multiple irreversible processes, the flux of each irreversible process j is the common result of forces of all irreversible processes. The coefficient L_{ij}, the influence of the force of the ith irreversible process on the flux of the jth irreversible process, is just equal to the coefficient L_{ji}, the influence of the force of the jth irreversible process on the flux of the ith irreversible process (i.e. $L_{ij} = L_{ji}$). In fact, the reciprocal relations are the influences, or cross effects, between individual irreversible processes in multi-process systems. These are also relations between coefficients related with thermodynamic coupling. Onsager (Yale University) won the 1968 Nobel Prize in Chemistry "for the discovery of the reciprocal relations bearing his name, which are fundamental for the thermodynamics of irreversible processes".

The similar reciprocal relation, or called by Onsager "Thomson reciprocal relation", associated with cross effects was first proposed by William Thomson (Lord Kelvin) as early as the end of the 19th century. [17] Thomson reciprocal relations were written in resistance form $R_{12} = R_{21}$. As Onsager said in 1931, [16]

> Thomson relation is generally accepted, because it has been confirmed within the limits of error of the best measurements.

As Onsager also said about the conductance form $L_{12} = L_{21}$, [16]

> Helmholtz has derived the reciprocal relation.

and

> the experiments confirm the result, $\cdots$ and the theorem is generally accepted.

That was Helmholtz's reciprocal relation $L_{12} = L_{21}$. [18] Onsager reciprocal relations were written in a more general form $L_{ij} = L_{ji}$. However, as Onsager said, his assumption was not to the requirements of thermodynamics. Experimental data had confirmed the approximation of Onsager reciprocal relations. [2] Onsager reciprocal relations are usually regarded

---

[29] http://nobelprize.org/nobel_prizes/lists/1968.html. Cited 18 Oct 2008

as belonging to the field of linear dissipative thermodynamics (previously called linear nonequilibrium thermodynamics or linear thermodynamics for irreversible processes).

Fig. 2.26 Ilya Prigogine (1917–2003)

In the 1950s and 1960s, Ilya Prigogine (Fig. 2.26)[30] of Universite Libre de Bruxelles (ULB), Belgium, deduced the entropy production minimization principle in the regime of linear dissipative thermodynamics on the basis of Onsager reciprocal relations. Prigogine's contributions were mainly in the nonlinear dissipative thermodynamics (previously called nonlinear nonequilibrium thermodynamics, or thermodynamics for nonlinear irreversible processes). He proposed the dual role of irreversible processes: as destroyers of order near equilibrium and as creators of order far from equilibrium. That was a qualitative description of thermodynamic coupling, which will be discussed later.

Prigogine proposed that irreversible processes might be the source of ordering, and from macroscopic view explained a large number of existing spatio-temporal organized structures. Since the creation and maintenance of organized nonequilibrium states rely on dissipative processes, they are called the theory of dissipative structures. [19, 20] In nonequilibrium systems, oscillating concentrations and geometrical concentration patterns can be a result of chemical reactions and diffusion. In order to simulate the chemical oscillation, a typical example of dissipative structures, Prigogine had to introduce "Brusselator", which was a "trimolecular model" of autocatalytic chemical kinetics, as pointed out by Prigogine and his co-workers themselves. Ilya Prigogine won the 1977 Nobel Prize in Chemistry "for his contributions to non-equilibrium thermodynamics, particularly the theory of dissipative structures".

These two Nobel Prizes terms in the 1960s and 1970s given for "irreversible processes" and "non-equilibrium thermodynamics" are contributions to the development of researches on complex systems. Onsager and Prigogine were outstanding scientists in thermodynamics with a deep fundamental knowledge of physics, mathematics and perhaps even chemistry. However, at that time most problems, or so-called "abnormal" phenomena of thermodynamics for complex multi-process systems appeared in biological or biochemical systems. That may be the objective background for a biochemist, Dean Burk, to clearly propose in 1931 an important concept of thermodynamic coupling, or coupled reactions. This was directly on the basis of the second law of thermodynamics without any other out-of-thermodynamic assumptions or models in his paper, "*The Reversibility of Coupled Reactions in Biological Systems and The Second Law of Thermodynamics*". [21]

After that in 1961, a British biochemist, Peter D. Mitchell (Fig. 2.27)[31], proposed a chemiosmotic coupling hypothesis for ATP biosynthesis. [22] After checkback and experimental confirmations over more than 10 years Mitchell at last won the 1978 Nobel Prize in chemistry "for his contribution to the understanding of biological energy transfer through

---

[30] http://nobelprize.org/nobel_prizes/lists/1977.html. Cited 18 Oct 2008

[31] http://nobelprize.org/nobel_prizes/lists/1978.html. Cited 18 Oct 2008

the formulation of the chemiosmotic theory". That was also conclusive evidence for thermodynamic coupling being the core of modern thermodynamics, and that should really be considered an important contribution in the development of modern thermodynamics. However, Mitchell chemiosmotic coupling theory merely becomes part of the central contents of biochemistry textbooks, and only few thermodynamics textbooks in physics and chemistry mention it. Chapter 4 of this book will make a more detailed discussion on it.

The development of the concept of thermodynamics coupling, i.e. the development of the second law of thermodynamics, made thermodynamic coupling the core of modern thermodynamics. However, such a development has meandered through the thermodynamic coupling model for the activated low-pressure diamond growth and so on in recent years.[23∼26] That will be discussed in detail in the following chapters of this book.

Fig. 2.27 Peter D. Mitchell (1920–1992)

References

1. Wang Z-X. Thermodynamics [M]. Beijing: High Education Press, 1955.
2. Kondepudi D, Prigogine I. Modern Thermodynamics — From Heat Engines to Dissipative Structures [M]. New York: John Wiley & Sons, 1998. http://www.wiley.com/. Cited 20 Apr 2008.
3. Mendoza E (ed). Reflextions on the Motive Force of Fire by Sadi Carnot and other papers on the Second Law of Thermodynamics by E. Clepeyron and R. Clausius [M]. 1977; Reflexions sur la puissance motrice du feu, et sur les machines propres d velopper cette puissance. Paris: 1924.
4. Clapeyron E. Journal de l'Ecole Polytechnique [J], 1834, 14, 153-190.
5. Fu X-C, Shen W-X, Yao T-Y. Physical Chemistry [M]. 4th edn. Beijing: High Education Press, 1990.
6. Kelvin W T. Transactions of the Edinburg Royal Society [J], 2 Jan 1849, XVI.
7. Clausius R. The Mechanical Theory of Heat [M]. Tr. by Browne W.R. London: MacMillan, 1879, 89.
8. Clausius R. Mechanical Theory of Heat [M], London: John van Voorst, 1867.
9. Kastler A. L'Oeuvre posthume de Sadi Carnot, in Sadi Carnot et l'Essor de la thermodynamique, Editor AN (ed) [M]. Paris: CNRS, 1974.
10. Conant JB (ed). Harvard Case Histories in Experimental Science, vol 1 [M]. Cambridge, MA: Harvard Univ. Press, 1957; Mason SF.A History of the Sciences [M]. New York: Collier Nooks, 1962.
11. Carathéodory C. Math Ann [J], 1909, 67, 355. See Born M. Phys Zeit [J], 1921, 22: 218, 249, 282; Z.-X. Wang, Thermodynamics [M]. Beijing: High Education Press, 1955, 403
12. Cápek V, Sheehan DP. Challenges to The Second Law of Thermodynamics: Theory and Experiment [M]. Berlin: Springer, 2004.
13. Gibbs JW. The Corrected Works of J.W. Gibbs, Vol I [M]. New York: Longmans, Green, 1931.
14. Nernst W. A New Heat Theorem [M]. New York: Free Press, 1969, 85.
15. Einstein A. The World As I See It [M]. translated by Harris A. New York: The Wisdom Library, a Division of Philosophical Library, 1949, 5.

16. Onsager L. Phys Rev [J], 1931, 37: 405; 1931, 38: 2265.
17. Thomson W (Lord Kelvin). Proc Roy Soc [C]. Edinburgh: 1854, 123; Collected Papers I: 237-242.
18. Von Helmholtz H. Wied Ann [J], 1876, 3: 201; Wiss Abh [J], 1: 840.
19. Prigogine I. Introduction to Thermodynamics of Irreversible Processes [M]. 3rd edn. New York: Interscience Publishers, John Wiley & Sons, 1967.
20. Li R-S. Nonequilibrium Thermodynamics and Dissipative Structure [M], Beijing: Tsinghua University Press, 1986.
21. Burk D. J Phys Chem [J], 1931, 35: 432.
22. Mitchell P. Science [J], 1979, 206: 1148-1159.
23. Wang J-T. Modern Thermodynamics — and a Whole View of Thermodynamics [M]. Shanghai: Fudan Univ. Press, 2005, 212 pages.
24. Wang J-T. Nonequilibrium Nondissipative Thermodynamics — with Application to Low-Pressure Diamond Synthesis [M]. Heidelberg: Springer, 2002: 254.
25. Wang J-T. Phase Diagrams of Stationary Nonequilibrium States — Thermodynamics for Low Pressure Diamond Growth from the Vapor Phase [M]. Beijing: Science Press, 2000: 212.
26. Wang J-T, Zhang DW, Liu Z-J. Thermodynamic Coupling Model for Low Pressure Diamond Growth from the Vapor Phase [M]. Beijing: Science Press, 1998, reprinted in 2000: 175.

Chapter 3
Fundamentals of Classical Thermodynamics

Abstract The development of science, technology and cognition of human beings usually follows a rule from simple to complex, from low level to high level. The development of the thermodynamics discipline also follows this rule. Classical thermodynamics deals with the change rule of relatively simple macroscopic systems. The development of modern thermodynamics has also occurred on the basis of the achievements of classical thermodynamics. In this chapter, simple systems such as single gas systems, even the ideal gas systems, are the main research objects for discussion. Classical thermodynamics should be used only for simple spontaneous-process systems, equilibrium systems and systems with reversible process(es). However, classical thermodynamics still has more general and wider meanings. It may be used for gas, liquid, solid, multiphase systems and chemical reaction systems. Some statements and mathematical expressions of classical thermodynamics in key places in this book are different from most thermodynamics textbooks.

3.1 Some Fundamental Concepts in Classical Thermodynamics

Thermodynamics is a part of science which does research on energy and energy transformation, and its research objects are the change and development of macroscopic systems consisting of a large number of particles. Classical thermodynamics, however, does research on simple systems [i.e. systems with only spontaneous irreversible process(es), equilibrium systems and systems with reversible process(es)], or as will be discussed further in the next chapter, classical thermodynamics is a part of thermodynamics for research on uncoupling systems.

3.1.1 System and Surroundings

For detailed researches in thermodynamics, an object with macroscopic size and a well-defined geometrical volume is divided from the objective world, and is called a thermodynamic system, macroscopic system, or system for short. The research object in thermodynamics is usually a practical body, so sometimes system is also called body. The boundary of a system is a surface, by which the system can be separated from the exterior related part of the whole objective world. The exterior related part is called surroundings of the system. Note that, surroundings are still a limited part of the whole objective world, so that the system together with its surroundings still comprises only a part of the objective world, and does not equal the universe. Such definitions of a system and its surroundings agree with our daily experiences, and from such definitions the so-called "*entropy of the universe*" will not be considered equal to the sum of entropy of a system and entropy of its surroundings. "*The fundamental laws of the universe*" was an incorrect conclusion. (p. 365)[1]

If there is no exchange of matter and energy between a system and its surroundings, then the system is called an isolated system. If there is only exchange of energy but no exchange of matter between a system and its surroundings, then the system is called a closed system. If there are exchanges of both energy and matter between a system and its surroundings, then the system is called an open system.

In any cases, a system and its surroundings may be separated by different kinds of walls, and these may be adiabatic wall, heat-conductible wall and rigid wall. The system separated by an adiabatic wall is called adiabatic system. If the system is separated by a rigid wall, then there is no volumetric work between the system and its surroundings. If the system is separated by a rigid heat-conductible wall, then there may be heat conduction but no volumetric work between the system and its surroundings.

3.1.2 Equilibrium State and Nonequilibrium State

The macroscopic performance at different times and under different influences of external surroundings is called macroscopic state or state. When a system under invariable external restricted conditions (such as a fixed boundary condition or a concentration restricted condition etc.), a macroscopic change may take place inside the system. At that time the system is in a nonequilibrium state. After a certain time, the system may reach an invariable state independent of time, which is called a stationary nonequilibrium state, or a stationary state for short. If a system is in a stationary nonequilibrium state, the internal macroscopic process may be still going on. If the system is an isolated system, i.e. there is no external restricted condition, then the system must develop into an invariable state without any macroscopic process. That is a special stationary state — an equilibrium state. If there is no external restricted condition, closed systems or open systems may also develop into equilibrium states. Once a system reaches its equilibrium state, there must be no macroscopic process inside the system. Note that equilibrium states, i.e. states without internal macroscopic processes, sometimes can also be called zero-process systems or no-process systems.

The macroscopic performance of a stationary nonequilibrium state is independent of time, but macroscopic process inside the system may still be going on. That is, only the total result of internal changes and external exchanges makes the macroscopic state of the system unchanged. Therefore, it is necessary to distinguish between stationary nonequilibrium state and equilibrium state. Meanwhile, it could be found that the most general state of a system is nonequilibrium state. When a system is in its nonequilibrium state, macroscopic internal state of the system usually changes with time continuously. Stationary nonequilibrium state may be regarded as a special state of nonequilibrium states, while equilibrium state may be regarded as a special state of stationary nonequilibrium states. Among nonequilibrium states, stationary nonequilibrium state is of special importance.

Some examples can illustrate this. For instance, both ends of a metallic rod are contacted with a hot heat-reservoir and a cold heat-reservoir, respectively, after a certain time temperature at each point of the metallic rod is of a certain fixed value. At that time the metallic rod is in a stationary state, but not an equilibrium state, because temperatures at different points of the metallic rod are different and there is a macroscopic constant heat flow through the metallic rod. Another example is a living body at a certain stage. Here, the system is of a relatively stable macroscopic state: continuous eating foods, exhausting the waste and metabolism taking place. Such a living body can also be regarded as a stationary state.

It should be pointed out that only equilibrium states may be practically calculated in classical thermodynamics. In nonequilibrium systems or irreversible processes the quantitative calculations in classical thermodynamics are limited only to their initial equilibrium states and their final equilibrium states. Therefore, sometimes the terms "thermodynamics for equilibrium states" and "thermodynamics for nonequilibrium states" can be found. It was hoped that the term "thermodynamics for equilibrium states" could be used instead of classical thermodynamics. However, meanings of these terms are not very clear, and these terms are easy to confuse with widely used terms of equilibrium thermodynamics and nonequilibrium thermodynamics. These terms can possibly be regarded as synonyms, so the terms "thermodynamics for equilibrium states" and "thermodynamics for nonequilibrium states" will not be recommended in this book and will not be used in other parts of this book.

3.1.3 State Variables and State Functions

Thermodynamic systems are composed of a large number of particles. Thermodynamics does not make research on microscopic states of systems. Even if a system is in its equilibrium state, microscopic states of the system still change with time ceaselessly. Therefore, thermodynamic macroscopic equilibrium states are the result of microscopic kinetic balance of a large number of particle motions. Particles inside the system are in incessant motion. If the number of particles is large enough (such as in magnitude of 6.023×10^{23}), it is already impossible to do research on the basis of Newton mechanics, but it obeys rules of statistics. The corresponding macroscopic states are impossible to describe by microscopic mechanic variables of particles, but should be described by some macroscopic variables. These variables are called state variables, or state functions. For instance, gas systems may be described by volume, pressure and temperature.

If in the whole system all physical and chemical characters are uniform, then the system is called a homogeneous system or single-phase system. If in an equilibrium system there are several parts of uniform characters, then the system is called a multi-phase system. The definition of phase is the uniform part of all physical and chemical characters, there is a clear boundary between phases under the appointed conditions, and the characters change suddenly at the boundary. For nonequilibrium systems, the definition of phase may have some extended meanings, which will be discussed later.

Among these state variables or state functions, temperature is a very special state variable or state function. Temperature itself may be directly measured, and may be decided through other state variables. State variables or state functions such as internal energy, entropy, enthalpy, free energy and Gibbs free energy are impossible to be measured directly.

When a system is in its equilibrium state, the characters of the system are only dependent on the state of the system, but independent of the history of the system. Characters of the system change with the change of its state. The quantity of change depends only on the initial and final states of the system, but is independent of path of the change. State variables or state functions may be divided into two groups: State variables or state functions, such as temperature, pressure, which are independent of the quantity of matter, belong to the first group, called intensive variables; while state variables or state functions such as volume, weight, internal energy, which depend on the quantity of matter, belong to the second group, called extensive variables. [2, 3] Intensive variables are of local characteristics but with no additivity property. That is, the value of these intensive variables measured at any local position in an equilibrium system will be the same as that measured in the whole system, but this does not allow taking the sum of local values as the value for the whole system. However, extensive variables are of additivity property, i.e. the sum of all local measured values is the value of the whole system. These different local characteristics are the basis of local equilibrium approximation, which must be introduced for treatment and understanding of nonequilibrium states. For instance, the values of temperature or pressure at different local positions may be different, and there may be no uniform temperature and pressure for the whole nonequilibrium system. However, while internal energy is of the additivity property and the value of different local parts may be different, the sum of internal energy values of all local parts is still the internal energy value for the whole system.

Thermodynamics cannot decide how many independent state parameters are needed for description of a system, but it has been confirmed by experiments that for a single-substance single-phase closed system without chemical reactions two intensive variables are usually needed to ensure other intensive variables, and that if the total quantity of the system is fixed, then extensive variables are ensured as well.

3.1.4 Reversible, Irreversible and Quasistatic Processes

The state change of a system with time is called macroscopic process or process. If the external surroundings of an equilibrium system are changed, then the equilibrium state of the system may be destroyed. At that time the system is in nonequilibrium state, and the system changes together with the external surroundings. Finally, the system will reach a new equilibrium state corresponding to the external surroundings. That is a process. In thermodynamics there is also an important concept of reversible process. In classical thermodynamics if a process starts from initial equilibrium state of a system, and during the process both the system and its surroundings may resume to their initial states, then the process is called reversible process. The definition of the concept of reversibility requires the system together with its surroundings being resumable. The nature of reversible process is that the same values with opposite directions for all heat and work interactions take place at the boundary. Therefore, when the system returns back to its initial state no historical trace is left in the surroundings. That is only an ideal process, which might be approached by quasistatic process. Quasistatic process should be composed of a series of processes, in which both the system and its surroundings are always in equilibrium states. Quasistatic process may help us to understand reversible process. Any practically occurring macroscopic process is irreversible process, i.e. when the system returns back to its initial state an historical trace must be left in the surroundings. Any spontaneous process taking place in isolated system must be irreversible process, and increases entropy of the system.

3.1.5 Spontaneous and Nonspontaneous Processes

In thermodynamics, processes may also be divided into spontaneous processes and nonspontaneous processes. A process occurring along such a direction that the process may automatically go on without other influences left is called spontaneous process; while a process taking place along a direction opposite to that of spontaneous process is called nonspontaneous process. Spontaneous process is irreversible process. Any nonspontaneous process taking place practically is also irreversible process. In macroscopic world, the direction of single spontaneous process is very clear. In particular, a spontaneous process in an isolated system has no influence left on surroundings. For instance, the direction of heat conduction from high temperature to low temperature, the direction of diffusion from high concentration to low concentration, gas expansion to vacuum, etc. are of very clear directionality. A lot of these macroscopic experiences of human beings become the basis of the second law of thermodynamics. After the establishment of the first law and the second law of thermodynamics, thermodynamics becomes an important fundamental discipline in modern science. The single-way directionality of single spontaneous processes is the embodiment of the nature, “arrow of time”. It should be noticed that characteristic directionality is that only for the system including only one spontaneous process.

3.2 Mathematical Expressions of Basic Laws of Thermodynamics

The definition of temperature is a numerical value presentation of hotness or coldness of an equilibrium material object or system. Temperature is a special concept of thermodynamics. For a nonequilibrium system temperature may not be a fixed value. Here the local equilibrium approximation must be used. The definitions of temperature and temperature measurement should also be presented by the zeroth law of thermodynamics (or called heat equilibrium law). That is, if two systems are in heat equilibrium with each other, then their temperatures must be equal. Here, “two systems are in heat equilibrium with each other” means that there is no macroscopic heat flow between these two heat-connected systems. If the law is presented by temperature T, then in equilibrium, $T_1 = T_2$. On the contrary, this law is still correct, i.e. if $T_1 = T_2$, then these two systems are also in heat equilibrium with each other. That is the condition of heat equilibrium. Thus, temperature of a system may be measured by another system used as a standard thermometer under the condition of heat equilibrium between them. If the scale of temperature has been selected, then the value of temperature may be obtained. The zeroth law of thermodynamics can also be described as follows: if system 1 is in heat equilibrium with system 2 and system 3 respectively at the same time, then system 2 and system 3 must be in heat equilibrium on contact with each other, i.e. if $T_1 = T_2$ and $T_1 = T_3$, then $T_2 = T_3$ must be correct. Note that, the scientific definition of temperature might be different from our feeling. For instance, a bulk of iron and a bulk of wood at room temperature actually have the same temperatures, but based on our feeling the bulk of iron would be colder. This is because the bulk of iron easily removes heat from our body. Therefore, the zeroth law of thermodynamics was obtained after recognition of the difference between heat and temperature. In spite of the early recognition of such a fundamental law its name as the zeroth law of thermodynamics was given only after the establishment of the first law, the second and the third law of thermodynamics for reparation of the completeness of the fundamental laws.

3.2.1 Expression of the First Law of Thermodynamics

In 1847, the mechanical equivalent of heat was proved by James Prescott Joule. That is, there is equivalence between heat and mechanical work: a certain amount of mechanical work, regardless of the particular methods of conversion, corresponds always to the same amount of heat (4.184 joules produce 1 calorie of heat). In this way, the mechanical equivalent of heat was later generalized into the law of conservation of energy, i.e. the energy sum of a system and its surroundings keeps constant. The total energy of a system and its surroundings may change from one form to another. Whenever a quantity of one kind of energy disappears an exactly equivalent amount of other kinds must be produced. In other words, energy is impossible to be created or to be destroyed in an isolated system. That is the written statement of the first law of thermodynamics.

The first law of thermodynamics has very important practical meanings. At the time many researchers tried to produce a perpetual motion machine, which would continuously produce mechanical work without supplying equivalent amount energy from other sources. That could be called the first kind of perpetual motion machine. That is, of course, against the first law of thermodynamics, so a brief written statement of the first law of thermodynamics is that the production of the first kind of perpetual motion machine is impossible. Up to now, nobody has found any fact contrary to this empirical rule from the daily activities of human beings. That is the most powerful demonstration for the first law of thermodynamics.

Mathematically, the first law of thermodynamics can be formulated for a closed system. That is, the change of internal energy of a system $\mathrm{d}U$ (positive for increase, and negative for decrease; this will be the same for other state functions below) equals the heat absorbed by the system δQ minus the work done by the system δW.

$$\mathrm{d}U = \delta Q - \delta W. \tag{3.1}$$

Here, the change of internal energy $\mathrm{d}U$ is in differential form, because its value depends only on the initial state and the final state of the system, but is independent of the path of the process. Heat absorbed by the system δQ and work done by the system to exteriors are quantities of process, and depends on the practical path of change, so there are some differences in expressions.

For an isolated system there is no exchange of heat and work between the system and its surroundings, and no exchange of matter either. Therefore, for an isolated system

$$(\mathrm{d}U)_{\mathrm{iso}} = 0. \tag{3.2}$$

That is, the internal energy of an isolated system always keeps constant.

After the first law of thermodynamics had been established, some still hoped to find other kinds of machines which could produce mechanical work by supplying an equivalent quantity of heat from a single heat reservoir. If that could be realized, one would be able to get such a machine which could do work from the ocean as a heat reservoir. In this way, all the machines in the world might work for hundreds of years with a negligible decrease in the temperature of the ocean. It would be the second kind of perpetual motion machines. In general, the second kind of perpetual motion machines would drive the original automatically running macroscopic process in nature backward to its starting point without any other influences left. The above-mentioned example was practically driven by the heat production of friction backward without any other influences left, so it is impossible. This suggestion is not contrary to the first law of thermodynamics, but it has never been realized. From a lot of failures another basic rule had been obtained as the second law of thermodynamics.

3.2.2 Expression of the Second Law of Thermodynamics

Carnot theorem was also regarded as one kind of written statement of the second law of thermodynamics. However, Carnot theorem summarized a large amount of macroscopic experiences of human beings in the 19th century before 1824, so Carnot theorem should only be regarded as one kind of written statement of the second law of classical thermodynamics. More detailed discussions will be continued in the book hereafter.

According to Carnot theorem, Kelvin introduced an absolute scale of temperature. In this way, the efficiency of a reversible heat engine could be written as a simple function of temperatures of the hot heat reservoir and the cold heat reservoir, and the efficiency will never be over 1. The zero point in the absolute scale of temperature is independent of the property of any material. The full import of the general principle originating in Carnot theorem was made by Clausius through the introduction of the concept of entropy, which was a new physical quantity similar to energy. However, both the introduction of entropy and the setup of mathematical expression of the second thermodynamic law were made by Clausius still on the basis of Carnot theorems for the simple systems, as discussed in Carnot cycle. So classical thermodynamics should be used for simple systems. That is the basic limitation of classical thermodynamics.

A general and brief written statement of the second law of thermodynamics is that it is impossible to construct the second kind of perpetual motion machine. Or more specifically, it is impossible to construct a machine functioning in cycles, which can drive the original automatically running macroscopic process backward to its starting point without any other influences left. For instance, such a machine could convert heat completely into an equivalent amount of work without producing other changes elsewhere. The term "functioning in cycles" is inserted to indicate that the machine must resume exactly to its original state at regular stages, so that it can operate continuously. In 1850, Rudolf Clausius said that heat cannot of itself, without the intervention of any external agency, pass from a colder to a hotter body. In 1851, Lord Kelvin stated that it is impossible to obtain work by cooling a body below the lowest temperature of the system. These statements were regarded as equivalent in nature for other statements of the second law of thermodynamics, but they were concerned individually with an example of a simple spontaneous-process system. The general written statement of the second law of thermodynamics should be that it is impossible to construct a machine, which might make any single spontaneous irreversible process going back without any influence left.

As emphasized above, both the first law and the second law of thermodynamics were summarized from a large amount of macroscopic activity experiences of human beings, and could not be deduced from any more fundamental law in other disciplines of science. With the development of such a discipline of science it may be necessary to refine some of the original contents of thermodynamics teaching courses. The research objects of both thermodynamics and statistical physics are groups of a large amount of particles, so in some cases these two courses had been combined together into one course. However, the tendency of gradually substituting thermodynamics with statistical physics should be prevented. Another tendency which says that the basic laws of thermodynamics were deduced from statistical physics should also be avoided. This is not true in the history of development. On the other hand, the establishment of statistical physics was based on a fundamental assumption, while the establishment of the basic laws of thermodynamics was without such assumption. On the basis of the fundamental assumption of statistical physics, the deduction of the basic laws of thermodynamics in statistical physics could only be regarded as a confirmation of the correctness and acceptability of the fundamental assumption of statistical physics. Based on today's practical situations, not only was the birth of thermodynamics earlier than statistical physics, but also some developments in thermodynamics still keep ahead of statistical physics, especially those in the field of thermodynamics for nonequilibrium systems or modern thermodynamics. All of new developments in modern thermodynamics provide a powerful apocalypse to statistical physics.

The mathematical expression of the second law of thermodynamics was first proposed in 1865 by Clausius in his famous article. (p. 328–329)[1] It has already been mentioned in Chapter 2, Brief History of Thermodynamics, but due to its importance Fig. 2.4 will be repeated here as Fig. 3.1 although the figure caption may be not completely the same. Some of Clausius' original words will also be repeated for further discussions, but mathematical expressions will be written in the present general forms as possible. Clausius said: (p. 328–329)[1]

> In order to express analytically the second fundamental theorem in the simplest manner, let us assume that the changes which the body suffers constitute a cyclical process, whereby the body returns finally to its initial condition. By dQ we will again understand an element of heat absorbed, and T shall denote the temperature, counted from the absolute zero, which the body has at the moment of absorption, or, if different parts of the body have different temperature, the temperature of the part which absorbs the heat element dQ. If we divide the thermal element by the corresponding absolute temperature and integrate the resulting differential expression over the whole cyclical process, then for the integral so formed the relation holds.

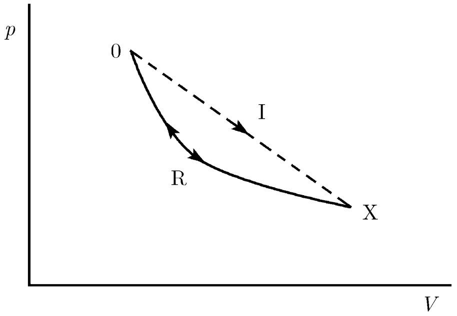


Fig. 3.1 Limitation of classical thermodynamics coming from the cyclical-process assumption and the assumption of initial and final equilibrium states

Clausius' "relation" could also be written in the present form, as Eq. (3.3).

$$\oint \frac{\mathrm{d}Q}{T} \leqslant 0. \tag{3.3}$$

He continuously said:

> in which the sign of equality is to be used in cases where all changes of which the cyclical process consists are reversible, whilst the sign < applies to cases where the changes occur in a non-reversible manner. * (* *d*Q is positive for heat absorbed by a changing body and is negative for heat given off by it.)

[Note: emphases were made by Clausius himself.]

Clausius emphasized the assumption of "cyclical process", and also emphasized that "the sign of equality" and "the sign <" were to be used for "reversible" and "non-reversible" processes, respectively. However, such an assumption of "cyclical process" was only included in his written statement and was not incarnated in his mathematical expression, Eq. (3.3).

Besides, Clausius made a definition of entropy of a system S. If the initial state and the final states are very close to each other, then the definition of entropy change can be written in differential form, i.e. entropy change of a system is the quotient of absorbed heat dividing by temperature "$\mathrm{d}S = \frac{\mathrm{d}Q}{T}$" in reversible processes. So in thermodynamics textbooks the formula was usually written as

$$\mathrm{d}S = \frac{\mathrm{d}Q_\mathrm{R}}{T}. \tag{3.4}$$

According to Carnot theorem, if two processes with the same initial and final equilibrium states to form a cyclical process are assumed, as shown in Fig. 3.1, then the heat absorbed by a system in a reversible process δQ_R is always bigger than the heat absorbed by a system in an irreversible process δQ. Therefore, the heat absorbed by a system in a practical process δQ divided by temperature T does not equal the entropy change of the system $\mathrm{d}S$, because any practical process is irreversible. Only for reversible processes in classical thermodynamics could the sign of equality be used. That is

$$\mathrm{d}S \geqslant \frac{\delta Q}{T} \text{ or } \mathrm{d}S - \frac{\delta Q}{T} \geqslant 0. \tag{3.5}$$

Equation (3.5) (i.e. Clausius inequality) is just as Eq. (2.21) discussed above. As also emphasized in Section 2.9, the process under consideration in classical thermodynamics must be of initial and final equilibrium states. If "Fig. 3.1, Limitation of classical thermodynamics coming from the cyclical-process assumption and the assumption of initial and final equilibrium states for any irreversible process", is written in mathematical expression, then $\mathrm{d}S_\mathrm{p} = \mathrm{d}S$ can be obtained. Here $\mathrm{d}S_\mathrm{p}$ and $\mathrm{d}S$ are entropy change of the process and entropy change of the system, respectively. That is a severe restriction of classical thermodynamics. Only a few examples of nonequilibrium systems or irreversible processes, which will be discussed in Section 3.6 , may satisfy such a requirement.

For isolated (or adiabatic) systems, because of no heat being absorbed by the system, the mathematical expression of the second law in classical thermodynamics can usually be simply written as

$$(\mathrm{d}S)_\mathrm{iso} \geqslant 0. \tag{3.6}$$

That is the same as Eq. (2.23), called the entropy increase principle.

If the limitation of classical thermodynamics $\mathrm{d}S_\mathrm{p} = \mathrm{d}S$ is included, then the classical mathematical expression of the second law can be more clearly written as

$$(\mathrm{d}S_\mathrm{p})_\mathrm{iso} = (\mathrm{d}S)_\mathrm{iso} \geqslant 0. \tag{3.7}$$

Here, $(\mathrm{d}S_\mathrm{p})_\mathrm{iso}$ is the entropy change due to the internal process in an isolated system, $(\mathrm{d}S)_\mathrm{iso}$ is the entropy change of the isolated system, and both of them are always with the same value and the same sign. [1] That is, in the framework of applicability of classical thermodynamics for isolated systems the entropy change of the internal process is just the entropy change of the system. Therefore, if the concerned object is inside the framework of applicability of classical thermodynamics, i.e. only simple spontaneous-process systems and equilibrium systems are under consideration, then the expression of Eq. (3.7) seemed to not be necessary. That might be a reason that except in some strict places (such as in Section 3.5 Criterion of Equilibrium, and discussions on classification of thermodynamics in other chapters), the form with only the state functions of systems (similar to Eq. (3.6) $(\mathrm{d}S)_\mathrm{iso} \geqslant 0$) will still be used in other parts of classical thermodynamics in this book.

The third law of thermodynamics is the impossibility of reaching zero degree in the absolute temperature scale (0 K), or the impossibility of cooling a body to 0 K by limited steps or procedures. According to the third law of thermodynamics, T will never be negative. In other words, the entropy of a pure crystalline substance may be taken to be zero at the absolute zero of thermodynamic temperature, i.e. 0 K. It may also be shown that the entropy of pure crystalline substances at absolute zero is not a function of pressure, i.e. $\left(\frac{\partial S}{\partial p}\right)_{T\to 0} = 0$. At temperatures above absolute zero, however, the entropy of a substance is a function of the pressure. Because of this the tabulated values of the absolute

---

[1] In the next chapter it will be known that the general entropy production form of $[\mathrm{d_i}S_\mathrm{p} = \mathrm{d_i}S \geqslant 0]$ should be used for classical thermodynamics.

entropy, relative to a value of zero at a temperature of absolute zero, are usually given at the standard reference pressure of 101.325 kPa. This reference state is usually denoted by the superscript (^{0}) on S, i.e. S^0. The absolute entropy of a substance can be evaluated by two distinct methods. One of these requires specific heat and latent heat data, while the other is statistical in origin and requires molecular information. The third law is more important for materials under very low temperature, so it will not be discussed in further detail in this book.

The total entropy of an isolated (or adiabatic) system increases irreversibly with time. In the next chapter we will talk about modern thermodynamics, in which entropy production of any system is never negative. Such irreversibility with time is rarely found in other fields of physics. For instance, in classical mechanics a system composed of n particles may be described by a phase space composed of n coordinate variables and n momentum variables. In such a phase space, trajectories started from different initial conditions will never cross each other. Different states of motion on a special trajectory are equivalent and reversible. For classical mechanic motions the concept of a four-dimensional space composed by positions and time may be used. The coordinate of time may also be positive or negative. However, for thermodynamic systems the change with time is usually in single direction. The total entropy of an isolated system changes always toward the direction to increase entropy; the change toward the opposite direction is impossible and the change is irreversible. For any simple spontaneous-process isolated (or adiabatic) system the total entropy change of the system does always agree with the entropy change of process. Both of them must increase. When the development of the total entropy change of an isolated (or adiabatic) system reaches its maximum, the system is in its equilibrium state, and such a system in its equilibrium state will never spontaneously go back to its initial state or any nonequilibrium state which it had already gone through. Nevertheless, different initial nonequilibrium states may reach the same equilibrium state.

3.3 Classical Equilibrium Thermodynamics

From the middle of the 19th century to the beginning of the 20th century, on the basis of mathematical expressions of the first and second laws of thermodynamics, a rather huge, strict and systematic thermodynamics discipline had been deduced, so thermodynamics developed at that time was often later called classical thermodynamics. In particular, equilibrium thermodynamics which it comprised was suitable to any equilibrium system. In equilibrium thermodynamics, both expressions of the first and second laws of thermodynamics are in their equality forms, which can be combined with each other to deduce a whole discipline of equilibrium thermodynamics. Here a lot of quantitative calculations for phase diagrams had been done. Therefore, equilibrium thermodynamics became a main part of classical thermodynamics. It was very easy to think that classical thermodynamics might be equivalent to equilibrium thermodynamics, or that equilibrium thermodynamics might represent the whole field of thermodynamics. However, it should be pointed out that the mathematical expression of the second law of thermodynamics consists of two parts: equality and inequality. In classical thermodynamics, the equality part was used for equilibrium systems or systems with only reversible processes, while the inequality part was only practically used for cyclical processes, including irreversible process(es), and for nonequilibrium systems with both initial and final equilibrium states as assumed by Clausius during the introduction of entropy and the mathematical expression of the second thermodynamic law. Therefore, classical thermodynamics, even in its much early stage, already included two parts of equilibrium thermodynamics and nonequilibrium thermodynamics. For classical equilibrium thermodynamics, the research objects are equilibrium systems or reversible processes, so the second law expression of classical equilibrium thermodynamics is an equality, $\mathrm{d}S = \frac{\delta Q}{T}$. Therefore, it is easy to combine this with the equality of the first law such as $\mathrm{d}U = \delta Q - \delta W$ for closed systems, and to deduce a whole systematic classical equilibrium thermodynamics. On the other hand, the second law for classical nonequilibrium thermodynamics is an inequality, $\mathrm{d}S > \frac{\delta Q}{T}$. Its research objects would be some nonequilibrium systems or irreversible processes. Due to the inequality expression of classical nonequilibrium thermodynamics it was difficult to make further quantitative calculation and deduction. Therefore, classical nonequilibrium thermodynamics can only calculate the initial and the final equilibrium states of a nonequilibrium system or an irreversible process, and compare the change value of their state function with experimental data to judge the nonequilibrium system or the irreversible process. In classical thermodynamics, there is no strict definition of state function for nonequilibrium states, or a local equilibrium approximation has to be introduced (which will be discussed in detail in the next chapter), so it is practically impossible to further treat nonequilibrium systems or irreversible processes by classical thermodynamics.

In this way, the development of nonequilibrium thermodynamics had been at a low ebb and had not attracted much attention for a long time.

It should be pointed out that the systematic development of classical equilibrium thermodynamics has also provided a firm foundation for nonequilibrium thermodynamics or even for modern thermodynamics. A lot of concepts, functions and relationships among state functions in nonequilibrium thermodynamics or modern thermodynamics are still based on equilibrium thermodynamics. Therefore, before a discussion of nonequilibrium thermodynamics and modern thermodynamics, some basic aspects of equilibrium thermodynamics and an overview of classical thermodynamics should be discussed first.

The basic parameters or state functions in thermodynamics are temperature T, pressure p, volume V, internal energy U and entropy S. For convenience of applications under different conditions, the enthalpy H was also introduced. Later, Hermann von Helmholtz and Josiah Willard Gibbs introduced two other state functions, i.e. Helmholtz free energy (simply called free energy or called Helmholtz isothermal isochoric free energy) F, and Gibbs free energy (or called Gibbs isothermal isobaric free energy) G, respectively. The definitions of these three derivative functions are

$$H = U + pV, \tag{3.8}$$

$$F = U - TS, \tag{3.9}$$

$$G = H - TS. \tag{3.10}$$

For remembering these definitions, the relationships among them can be illustrated as shown in Fig. 3.2.

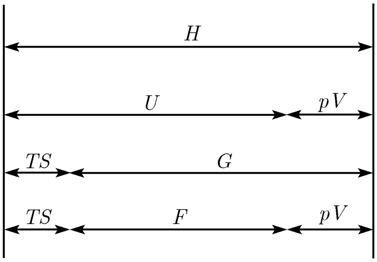


Fig. 3.2 Relations among thermodynamic functions

For closed systems, the mathematical expression of the first law is $\mathrm{d}U = \delta Q - \delta W$. To distinguish between expansion work $p\mathrm{d}V$ and other forms of work $\delta W'$, Eq. (3.1) can be rewritten as

$$\mathrm{d}U = \delta Q - p\mathrm{d}V - \delta W'. \tag{3.11}$$

According to the definition of H, this equation can be changed into

$$\mathrm{d}H = \delta Q + V\mathrm{d}p - \delta W'. \tag{3.12}$$

The physical meanings of U and H are clearly shown in Eq. (3.11) and Eq. (3.12). In isochoric processes with no work done by the system, the increase of the internal energy $\mathrm{d}U$ equals the heat absorbed by the system δQ during the process. In isobaric processes with only expansion work, the increase of enthalpy $\mathrm{d}H$ equals the heat absorbed by the system δQ during the process.

For an equilibrium system or for a system with only reversible processes, the mathematical expression of the second law is

$$T\mathrm{d}S = \delta Q. \tag{3.13}$$

Therefore, according to the definitions of F and G the combined mathematical expression for the first law and the second law of thermodynamics can be written as follows.

$$\mathrm{d}F = -S\mathrm{d}T - p\mathrm{d}V - \delta W' \tag{3.14}$$

and

$$\mathrm{d}G = -S\mathrm{d}T + V\mathrm{d}p - \delta W'. \tag{3.15}$$

The physical meanings of F and G can be found in Eq. (3.14) and Eq. (3.15), i.e. in isothermal and isochoric processes the decrease of the Helmholtz free energy of the system $-\mathrm{d}F$ equals the work done by the system in a reversible process $\delta W'$, and in isothermal and isobaric processes the decrease of the Gibbs free energy of the system $-\mathrm{d}G$ equals the nonexpansion work done by the system in a reversible process $\delta W'$.

For closed systems during reversible processes with only expansion work, the set of Eq. (3.11), Eq. (3.12), Eq. (3.14) and Eq. (3.15) can be simplified as follows.

$$\mathrm{d}U = T\mathrm{d}S - p\mathrm{d}V, \tag{3.16}$$

$$\mathrm{d}H = T\mathrm{d}S + V\mathrm{d}p, \tag{3.17}$$

$$\mathrm{d}F = -S\mathrm{d}T - p\mathrm{d}V, \tag{3.18}$$

$$\mathrm{d}G = -S\mathrm{d}T + V\mathrm{d}p. \tag{3.19}$$

Meanwhile, it may be found that all of U, H, F and G are in dimension of energy. In practice, all of them are criterion factors to decide the change direction by energy from high to low under different conditions, and the maximum and limitation of released heat or nonexpansion work.

Equation (3.16)–Eq. (3.19) are expressions of complete differentials, so a series of useful partial differential equations can be written as follows.

$$T = \left(\frac{\partial U}{\partial S}\right)_V = \left(\frac{\partial H}{\partial S}\right)_p, \tag{3.20}$$

$$p = -\left(\frac{\partial U}{\partial V}\right)_S = -\left(\frac{\partial F}{\partial V}\right)_T, \tag{3.21}$$

$$V = \left(\frac{\partial H}{\partial p}\right)_S = \left(\frac{\partial G}{\partial p}\right)_T, \tag{3.22}$$

$$S = -\left(\frac{\partial F}{\partial T}\right)_V = -\left(\frac{\partial G}{\partial T}\right)_p. \tag{3.23}$$

According to the properties of complete differentials of state functions, from Eq. (3.20) to Eq. (3.23) we may obtain the following equations.

$$\left(\frac{\partial T}{\partial V}\right)_S = -\left(\frac{\partial p}{\partial S}\right)_V, \tag{3.24}$$

$$\left(\frac{\partial T}{\partial p}\right)_S = \left(\frac{\partial V}{\partial S}\right)_p, \tag{3.25}$$

$$\left(\frac{\partial S}{\partial V}\right)_T = \left(\frac{\partial p}{\partial T}\right)_V, \tag{3.26}$$

$$\left(\frac{\partial S}{\partial p}\right)_T = -\left(\frac{\partial V}{\partial T}\right)_p. \tag{3.27}$$

The above four equations are usually called the Maxwell relationships, and they show the relationship among certain thermodynamic functions in homogeneous equilibrium systems. These equations are useful for getting some partial differentials, which are difficult to measure, through the terms of other easily measurable partial differentials.

As discussed in Chapter 2, the most difficult thing may be how to find the physical meaning of entropy among all the thermodynamic functions. It seems very difficult to find a physical quantity directly corresponding to the entropy. Roughly speaking, the value of the entropy may be regarded as a measure of the microscopic "randomness" or "disorder" in a given system. All spontaneous processes in an isolated system are accompanied by an increase of entropy and therefore lead to an increase in the random distribution or disorder. For an ideal gas consisting of simple ball-type particles the value of the entropy is quantitatively proportional to the logarithm of the microscopic thermodynamic probability of the system, i.e. $S = k\ln\Omega$. k is called the Boltzmann constant. Classical expression of the second law of thermodynamics is the entropy increase principle, that is the microscopic state of the system spontaneously tends to the most probable distribution state with the maximum thermodynamic probability.

In 1869, M.F. Massieu demonstrated that after suitably selecting independent variables, the equilibrium properties of a homogeneous system can be decided by one thermodynamic function which is called the characteristic function.[4] For instance, U is the characteristic function with respect to S and V. From the basic Eq. (3.16), i.e. $\mathrm{d}U = T\mathrm{d}S - p\mathrm{d}V$, $T = \left(\frac{\partial U}{\partial S}\right)_V$ in the Eq. (3.20) and $p = -\left(\frac{\partial U}{\partial V}\right)_S$ in the Eq. (3.21) can be obtained. If the function expression of internal energy U with respect to independent variables of S and V is known, after a first-order differential the function expression of T and p with respect to S and V may be obtained. After eliminating S from these two equations of T and p the state equation of the system will be obtained. If V is eliminated from these two equations of T and p, the function expression of entropy S with respect to T and p will be obtained. By substitution of these obtained S and V into the expression of the internal energy U, the expression equation of internal energy with respect to T and p can be obtained. Therefore, if the function expression of internal energy U with respect to S and V is known, then the equilibrium properties of the system can be decided by differentials. That is, U is the characteristic function with respect to independent variables of S and V. Be careful, the characteristic function is related with the selection of independent variables. If the independent variables of (S, p), (T, V) or (T, p) are selected, then the corresponding characteristic functions will be H, F or G, respectively.

An English sentence "*Good Physicists Have Studied under Very Fine Teacher*" with two arrows has been recommended as a mnemonic to be used by physics students to remember the Maxwell relations in thermodynamics, as shown in Fig. 3.3.[2] Here, the first letter of each word represents a main state function of thermodynamics. Comparison between the figure and the series of equations mentioned above enables one to see how to use it.

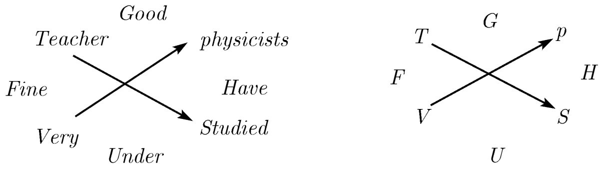


Fig. 3.3 An English sentence for remembering the relationships among thermodynamic functions

For instance, the four parameters G, H, U and F with the dimension of energy are located at the top, right, bottom and left, respectively. The conditions corresponding to them are located nearby. The state function G is located between T and p; it means that the state function G is the characteristic function with respect to independent variables of T and p. That is, the function G is a criterion for the spontaneous reaction direction at constant temperature and pressure. Similarly, F is located between T and V; it means that the function F is the characteristic function with respect to independent variables of T and V. That is,

---

[2] http://www.magicmnemonic.com/memory/examples32.html. Cited 23 Apr 2008

the state function F is a criterion for the spontaneous reaction direction at constant temperature and volume, and so on. All of these agree with the criteria mentioned in the next paragraph. Arrows in this figure are used for remembering signs of positive or negative. Gibbs and others in the 1870s and later successfully developed equilibrium thermodynamics. Except for a small part of content belonging to nonequilibrium thermodynamics, the discussions on equilibrium thermodynamics will still be continued in this chapter.

3.4 Classical Nonequilibrium Thermodynamics

All spontaneous processes are irreversible. From the inequality of the second law of thermodynamics in classical thermodynamics, the entropy of any isolated (or adiabatic) system with an irreversible process does always increase, i.e.

$$(\mathrm{d}S)_{\mathrm{iso}} > 0. \tag{3.28}$$

For closed systems

$$T\mathrm{d}S > \delta Q, \tag{3.29}$$

where δQ is the heat absorbed by the system in a practical process (i.e. an irreversible process), and T is the temperature of surroundings. The corresponding combination expression of the first law and the second law will be

$$\mathrm{d}U < T\mathrm{d}S - p\mathrm{d}V - \delta W', \tag{3.30}$$

$$\mathrm{d}H < T\mathrm{d}S + V\mathrm{d}p - \delta W', \tag{3.31}$$

$$\mathrm{d}F < -S\mathrm{d}T - p\mathrm{d}V - \delta W', \tag{3.32}$$

$$\mathrm{d}G < -S\mathrm{d}T + V\mathrm{d}p - \delta W'. \tag{3.33}$$

For closed systems with only expansion work, these formulae can be used as criteria for the direction of a single spontaneous process in a given system. The suitable conditions for each formula are different. For instance, the criterion for the direction of a single spontaneous process under isothermal and isochoric conditions should be $\mathrm{d}F < 0$, while the criterion for the direction of a single spontaneous process under isothermal and isobaric conditions should be $\mathrm{d}G < 0$. That is,

$$(\mathrm{d}U)_{S,V} < 0, \tag{3.34}$$

$$(\mathrm{d}H)_{S,p} < 0, \tag{3.35}$$

$$(\mathrm{d}F)_{T,V} < 0, \tag{3.36}$$

$$(\mathrm{d}G)_{T,p} < 0. \tag{3.37}$$

Because these equations are all inequalities, which can only be used as a criteria for the direction of a single spontaneous process, classical nonequilibrium thermodynamics was difficult for making any quantitative calculation and other further progress. Therefore, classical equilibrium thermodynamics had been regarded as the whole of classical thermodynamics or even as the whole of thermodynamics. That was a problem for the traditional point of view of thermodynamics, which led to some mistakes or misunderstandings.

For example, there had been a misunderstanding that diamond synthesis from graphite at low pressure had been regarded as thermodynamically impossible. In fact, the conclusion was an unsuitable statement coming from equilibrium thermodynamics, which is only suitable for equilibrium systems. The correct thermodynamic statement should be that diamond could not spontaneously form from graphite at low pressure, or that under low-pressure equilibrium conditions the transformation from graphite to diamond is impossible. The word "spontaneously" or "equilibrium" in the statement should never be omitted. That is to say, in the case of supplying external energy (i.e. under "nonspontaneous" or "nonequilibrium" conditions) the diamond synthesis from graphite at low pressure is not contrary to the fundamental principles of thermodynamics.[5~10] Activated low-pressure diamond synthesis was achieved by former Soviet Union scientists in about 1970, but the new technique was generally accepted only after the middle of the 1980s. The reason for this delay was due to the unsuitable or incomplete conclusions of classical equilibrium thermodynamics and the out-of-date traditional concept of thermodynamics. In fact, there are only three general statements about impossibilities in thermodynamics, i.e. the first, second and third laws of thermodynamics. That is, the first kind of perpetual motion machine, the second kind of perpetual motion machine and the absolute zero of temperature are impossible to practically achieve. All other impossibilities with general thermodynamic meanings in traditional concepts should be abandoned. In other words, all other impossibilities may be valid only under special conditions. If the conditions are changed they may become possible.

3.5 Criterion of Equilibrium

Since 1865, there has often been a common paradox in thermodynamics textbooks. On the one hand, the equality of the second law of thermodynamics [i.e. the first order differential of state function equals zero, such as $(\mathrm{d}S)_{\mathrm{iso}} = 0$] was regarded as the correspondence of an equilibrium system or a system including only reversible process without limitations.[3, 11] On the other hand, if the "criterion of equilibrium" was discussed in detail then it was often clearly pointed out that whether the first order differential of state function of a system equals zero is not a criterion of "equilibrium" for a system or "reversibility" for a process. This was the case in a textbook, "*Thermodynamics*", written by Z.-X. Wang as early as 1955. It was clearly pointed out that (p.152)[12]

> If the change of a system occurs under isolated conditions, the necessary and sufficient condition of thermodynamic equilibrium is the maximum of entropy.

and (p.158)[12]

> If the sufficient condition of equilibrium is needed, the second order differential of the related thermodynamics function must be discussed.

Strictly speaking, in classical thermodynamics if the entropy increase principle was used as the criterion of equilibrium for isolated (or adiabatic) systems, then it should be written in the following forms similar to Eq. (3.7).

1. For reversible process or equilibrium system

$$(\mathrm{d}S_{\mathrm{p}})_{\mathrm{iso}} = (\mathrm{d}S)_{\mathrm{iso}} = 0. \tag{3.38}$$

2. For irreversible process or nonequilibrium system

$$(\mathrm{d}S_{\mathrm{p}})_{\mathrm{iso}} = (\mathrm{d}S)_{\mathrm{iso}} > 0. \tag{3.39}$$

3. Directly, the entropy increase principle was written in a combined form like Eq. (3.7), i.e. entropy change due to process $(\mathrm{d}S_{\mathrm{p}})_{\mathrm{iso}}$ and entropy change of a system $(\mathrm{d}S)_{\mathrm{iso}}$ in classical thermodynamics are always with the same sign and the same value.

$$(\mathrm{d}S_{\mathrm{p}})_{\mathrm{iso}} = (\mathrm{d}S)_{\mathrm{iso}} \geqslant 0. \tag{3.40}$$

Here, the equality corresponds to an isolated system with reversible process or an isolated equilibrium system, while the inequality corresponds to an isolated system with irreversible process or an isolated nonequilibrium system.

It is well known that without external influences a system develops spontaneously toward its equilibrium state. In an isolated (or adiabatic) system, any single spontaneous process must increase entropy of the system. Finally, the system reaches an equilibrium state, with a maximum of entropy. Therefore, the criterion (the necessary and sufficient conditions) of an equilibrium state for an isolated (or adiabatic) system is a maximum of entropy for any possible change. That is, to reach the necessary and sufficient conditions of equilibrium the second order differential of state function of the system must be discussed.[5, 12, 13] In mathematics, that is $\delta S = 0$ together with $\delta^2 S < 0$. Note that if an isolated system has reached its equilibrium state there is no macroscopic process. Therefore, "any possible change" here means the phenomenon of fluctuation of microscopic particles, so the corresponding first order and second order differentials are presented by signs of variation.

At the same time it should be noted that if "any possible change" is concerned, then the system is not in equilibrium. In spite of a very small region close to equilibrium may be concerned, for the discussion on extremum the state function must be continuous and differentiable in this region, i.e., the local equilibrium approximation (see the next chapter) has been introduced. In practice, the often-used conditions are constant temperature and pressure conditions, so the Gibbs free energy, the function of G, is the most generally adopted form of criterion for a close system without nonexpansion work done. Similar to Eq. (3.7), the following two forms may be written.

1. For reversible process or equilibrium system

$$(\mathrm{d}G_{\mathrm{p}})_{T,p} = (\mathrm{d}G)_{T,p} = 0. \tag{3.41}$$

2. For irreversible process or nonequilibrium system

$$(\mathrm{d}G_{\mathrm{p}})_{T,p} = (\mathrm{d}G)_{T,p} < 0. \tag{3.42}$$

3. Directly, the Gibbs free energy decrease principle was written in a combined form like Eq. (3.7), i.e. Gibbs free energy change due to a process $(\mathrm{d}G_\mathrm{p})_{T,p}$ and Gibbs free energy change of a system $(\mathrm{d}G)_{T,p}$ in classical thermodynamics are always with the same sign and the same value.

$$(\mathrm{d}G_\mathrm{p})_{T,p} = (\mathrm{d}G)_{T,p} \leqslant 0. \tag{3.43}$$

Here, the equality corresponds to an isothermal isobaric system with reversible process or an equilibrium system, while the inequality corresponds to an isothermal isobaric system with irreversible process or an isothermal isobaric nonequilibrium system. In general, the second order differential of state function should be under consideration, i.e. for an isothermal isobaric system the necessary and sufficient conditions of equilibrium (the criterion of minimum of Gibbs free energy) is $\delta G = 0$ together with $\delta^2 G > 0$. The criterion of equilibrium for an isothermal isochoric system is $\delta F = 0$ together with $\delta^2 F > 0$, as shown in Table 3.1. [5, 13]

Table 3.1 Criteria of Equilibrium

Independent Variable	Characteristic Function	Criteria of Equilibrium	Applicability
U, V	S	$\delta S = 0$, $\delta^2 S < 0$	Isolated (or adiabatic) systems
T, V	F	$\delta F = 0$, $\delta^2 F > 0$	Isothermal isochoric systems in thermal equilibrium
T, p	G	$\delta G = 0$, $\delta^2 G > 0$	Isothermal isobaric systems in thermal equilibrium

Therefore, phase diagrams calculated on the basis of the Gibbs free energy minimization principle must be equilibrium phase diagrams, but phase diagrams calculated only on the basis of $(\mathrm{d}G)_{T,p} = 0$ for a multiprocess system may be equilibrium phase diagrams and may be nonequilibrium phase diagrams, for which discussions will be made in the following chapters.

3.6 Calculation of Entropy Changes

Entropy is a state function in thermodynamics. If both the initial state and the final state are in equilibrium, then entropy change of the system must be independent of the path through which the real process takes place. Therefore, before calculation one must first design a reversible process. Some examples will be discussed as follows. (p.112) [3]

For isothermal processes, we can use the following equation

$$\Delta S = \frac{Q_\mathrm{R}}{T}. \tag{3.44}$$

Example 1 Under the conditions of 373.2 K and the standard pressure, 1 mol H_2O absorbs the heat of 40,620 J during its isothermal isobaric evaporation. Such a process may be further treated as a reversible process, so we may obtain

$$\Delta S = \frac{Q_{\mathrm{R}}}{T} = \frac{40,620\,\mathrm{J}}{373.2\mathrm{K}} = 108.8\ \mathrm{J\ K^{-1}}. \tag{3.45}$$

Example 2 1 mol ideal gas isothermally expands to be a volume of 10 times of its original volume. Such a process may be realized through a reversible process or through an irreversible expansion to vacuum, as shown in Fig. 3.4. The entropy changes in these processes must be the same because both their initial states and their final states are the same. Both the entropy change of the reversible process or the irreversible expansion to vacuum are

$$\begin{aligned}(\Delta S)_{\mathrm{syst}} &= \frac{Q_{\mathrm{R}}}{T} = \frac{W_{\mathrm{max}}}{T} \qquad (3.46)\\ &= \frac{\int_{V_1}^{V_2} p\mathrm{d}V}{T} = nR\ln\frac{10}{1} = 19.14\ \mathrm{J\ K^{-1}}.\end{aligned}$$

Based on Clausius inequality $\Delta S - \sum\left(\frac{\delta Q}{T}\right) \geqslant 0$, the entropy change of its surroundings during the reversible isothermal expansion is -19.14 J K^{-1}, but the entropy change of its surroundings during the irreversible expansion to vacuum is zero. Therefore, during the reversible isothermal expansion the entropy change of the isolated system is $\Delta S_{\mathrm{iso}} = \Delta S_{\mathrm{syst}} + \Delta S_{\mathrm{surr}} = 0$, while during the irreversible expansion to vacuum the entropy change of the isolated system is $\Delta S_{\mathrm{iso}} = (\Delta S_{\mathrm{syst}} + \Delta S_{\mathrm{surr}}) = 19.14$ J K^{-1}. That agrees with the entropy increase principle.

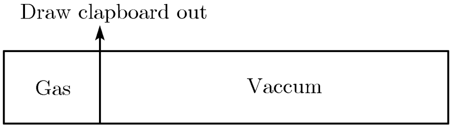


Fig. 3.4 An irreversible process of gas expansion to vacuum

Example 3 Under the standard conditions (298.2 K and 1013.3 kPa), 0.5 mol ideal gas A mixes with 0.5 mol ideal gas B to form a homogeneous gas mixture of A–B, as shown in Fig. 3.5.

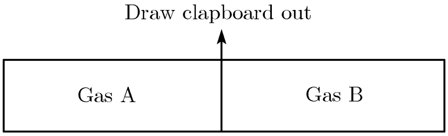


Fig. 3.5 An irreversible process of gas mixing

The entropy change of the mixing process $\Delta S_{\mathrm{mixing}}$ consists of two parts of ΔS_{A} and ΔS_{B}. ΔS_{A} equals the entropy change of the isothermal expansion of gas A from 11.2 dm^3 to 22.4 dm^3, so

$$\Delta S_{\mathrm{A}} = 0.5R\ln\frac{V_{\mathrm{A}}+V_{\mathrm{B}}}{V_{\mathrm{A}}} = 0.5R\ln\frac{V}{V_{\mathrm{A}}} = 0.5R\ln\frac{22.4}{11.2}. \tag{3.47}$$

The entropy change for gas B is the same.

$$\Delta S_{\mathrm{B}} = 0.5R\ln\frac{22.4}{11.2}. \tag{3.48}$$

Therefore, the entropy change of gas mixing equals

$$\Delta S_{\mathrm{mixing}} = \Delta S_{\mathrm{A}} + \Delta S_{\mathrm{B}} = R\ln\frac{22.4}{11.2}. \tag{3.49}$$

If initial pressures of all gases before mixing are the same, and the final pressure of the gas mixture is also the same, then the entropy change of mixing $\Delta S_{\mathrm{mixing}}$ may be written in the following general form:

$$\begin{aligned}\Delta S_{\mathrm{mixing}} &= n_{\mathrm{A}}R\ln\frac{V_{\mathrm{A}}+V_{\mathrm{B}}}{V_{\mathrm{A}}} + n_{\mathrm{B}}R\ln\frac{V_{\mathrm{A}}+V_{\mathrm{B}}}{V_{\mathrm{B}}} \\ &= -n_{\mathrm{A}}\ln x_{\mathrm{A}} - n_{\mathrm{B}}\ln x_{\mathrm{B}}.\end{aligned} \tag{3.50}$$

Here, n and x are molar numbers and molar fractions of individual gases, respectively.

For nonisothermal processes, the heat capacity should be under consideration during the calculation of the entropy change. The definition of heat capacity is

$$\delta Q = C\mathrm{d}T. \tag{3.51}$$

Therefore, the entropy change of the system due to the temperature change is

$$\mathrm{d}S = C\frac{\mathrm{d}T}{T}. \tag{3.52}$$

C_V is the isochoric heat capacity. The entropy change for an isochoric process is

$$\mathrm{d}S = C_V\frac{\mathrm{d}T}{T} \text{ and } \Delta S = \int C_V\frac{\mathrm{d}T}{T}. \tag{3.53}$$

C_p is the isobaric heat capacity. The entropy change for an isobaric process is

$$\mathrm{d}S = C_p\frac{\mathrm{d}T}{T} \text{ and } \Delta S = \int C_p\frac{\mathrm{d}T}{T}. \tag{3.54}$$

Example 4 The entropy change of 1 mol ideal gas for a process from the initial state of A (p_1, V_1, T_1) to the final state of B (p_2, V_2, T_2) may be calculated through two different paths of reversible processes.

The first path consists of two steps: The first step is from A (p_1, V_1, T_1) to C (p, V_2, T_1) and the second step is from C (p, V_2, T_1) to B (p_2, V_2, T_2), as shown in Fig. 3.6.

$$\Delta S = \Delta S_1 + \Delta S_2 = nR\ln\frac{V_2}{V_1} + \int_{T_1}^{T_2} C_V\frac{\mathrm{d}T}{T}. \tag{3.55}$$

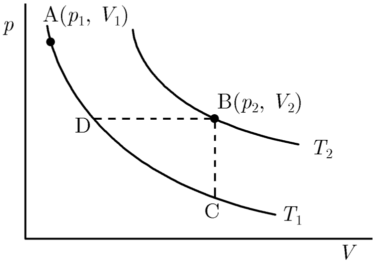


Fig. 3.6 Calculation of entropy change through two different paths of reversible processes.

The second path also consists of two steps: The first step is from A (p_1, V_1, T_1) to D (p_2, V, T_1) and the second step is from D (p_2, V, T_1) to B (p_2, V_2, T_2), as shown in Fig. 3.6.

$$\Delta S = \Delta S_{1'} + \Delta S_{2'} = nR\ln\frac{p_1}{p_2} + \int_{T_1}^{T_2} C_p\frac{\mathrm{d}T}{T}. \tag{3.56}$$

The calculated entropy change through two paths of reversible processes must be the same. Based on the relation $C_p - C_V = nR$, we can get

$$\begin{aligned}\Delta S_1 + \Delta S_2 &= nR\ln\frac{V_2}{V_1} + \int_{T_1}^{T_2} C_V\frac{\mathrm{d}T}{T} \qquad (3.57)\\ &= nR\ln\frac{p_1T_2}{p_2T_1} + \int_{T_1}^{T_2} C_p\frac{\mathrm{d}T}{T} - nR\ln\frac{T_2}{T_1}\\ &= nR\ln\frac{p_1}{p_2} + \int_{T_1}^{T_2} C_p\frac{\mathrm{d}T}{T} = \Delta S_{1'} + \Delta S_{2'}.\end{aligned}$$

From the above discussion, it has been confirmed that the entropy change is independent of the path through which the process takes place.

For other nonisothermal processes the calculation of entropy change will be discussed in the following examples.

Example 5 1 mol silver ingot is heated in vacuum from 273.2 K up to 303.2 K. It is known that the molar isochoric heat capacity of silver in such a temperature range is 24.48 J K^{-1} mol^{-1}. The expansive work in vacuum should not be under consideration, so this process may be treated as an isochoric one. The entropy change of this process is

$$\begin{aligned}\Delta S &= C_V\ln\frac{T_2}{T_1} \qquad (3.58)\\ &= (1\ \text{mol}) \times \left(24.48\ \text{J K}^{-1}\ \text{mol}^{-1}\right) \times \ln\frac{303.2}{273.2}\\ &= 2.531\ \text{J K}^{-1}.\end{aligned}$$

Example 6 25 dm^3 volume of 2 mol ideal gas is heated from 300 K up to 600 K, and after heating its volume becomes 100 dm^3. The known isochoric heat capacity is

$C_V = (19.37 + 3.39 \times 10^{-3}T)$ J K^{-1} mol^{-1}. The entropy change will be

$$\begin{aligned}\Delta S &= (2\,\text{mol}) \times \left(8.314\,\text{J K}^{-1}\,\text{mol}^{-1}\right) \times \ln\frac{100}{25} \qquad (3.59)\\ &\quad + (2\,\text{mol}) \times \int_{300}^{600}\left(\frac{19.37 + 3.39 \times 10^{-3}T}{T}\right)\mathrm{d}T\,\left(\text{J K}^{-1}\,\text{mol}^{-1}\right)\\ &= 51.96\,\text{J K}^{-1}.\end{aligned}$$

Example 7 1 mol liquid benzene is frozen at T = 268.2 K (-5 ^{o}C) and under the standard pressure. Such a process is an exothermic process with a heat release of 9,874 J. It is known that the melting point of benzene is 278.7 K (5.5 ^{o}C), the standard melting heat is ΔH = 9,916 J mol^{-1}, the molar isobaric heat capacity of liquid benzene is C_p(l) = 126.8 J K^{-1} mol^{-1}, and the molar isobaric heat capacity of solid benzene is C_p(s) = 122.6 J K^{-1} mol^{-1}. This process is a supercooling solidification so it must be an irreversible process, as shown in Fig. 3.7.

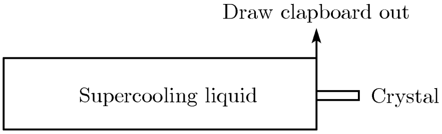


Fig. 3.7 An irreversible process of supercooling solidification

For calculation of the system entropy change ΔS, one must first design a reversible process to realize such a state change. Such a reversible may be divided into the following three steps: the first step is heating liquid benzene from 268.2 K to 278.7 K; the second step is a reversible solidification at 278.7 K; and the third step is cooling solid benzene from 278.7 K down to 268.2 K. The entropy changes of these reversible processes are represented by ΔS_1, ΔS_2 and ΔS_3, respectively. All of these entropy changes may be calculated individually as follows.

$$\begin{aligned}\Delta S_1 &= \left(\int_{268.2}^{278.7} 1 \times 126.8\frac{\mathrm{d}T}{T}\right)(\text{mol})\left(\text{JK}^{-1}\text{mol}^{-1}\right) \qquad (3.60)\\ &= 4.818\,\text{JK}^{-1},\\ \Delta S_2 &= \frac{Q_\text{R}}{T} = (1\,\text{mol}) \times \left(\frac{-9916}{278.7}\text{JK}^{-1}\text{mol}^{-1}\right) \qquad (3.61)\\ &= -35.58\,\text{JK}^{-1},\\ \Delta S_3 &= \left(\int_{278.7}^{268.2} 1 \times 122.6\frac{\mathrm{d}T}{T}\right)(\text{mol})\left(\text{JK}^{-1}\text{mol}^{-1}\right) \qquad (3.62)\\ &= -4.66\,\text{JK}^{-1}.\end{aligned}$$

Therefore,

$$\Delta S = \Delta S_1 + \Delta S_2 + \Delta S_3 = (4.818 - 35.58 - 4.66)\,\mathrm{JK}^{-1}. \tag{3.63}$$

It should be emphasized that the system entropy change is not a thermodynamic criterion for an open system or a close system. For instance, the above discussed supercooling solidification is a spontaneous irreversible process, but the system entropy change is of a negative value. Entropy change of a system may only be a thermodynamic criterion for an isolated system or an adiabatic system. For such a supercooling solidification the entropy change of surroundings must be under consideration as well. The surroundings may be regarded as a huge heat reservoir at 268.2 K. The heat absorption by a heat reservoir at a constant temperature, i.e. by surroundings, may be treated as a reversible process. Therefore, the entropy change of surroundings is

$$\Delta S_{\mathrm{surr}} = \frac{-\Delta H}{268.2\mathrm{K}} = \frac{9874\,\mathrm{J}}{268.2\mathrm{K}} = 36.82\,\mathrm{J\,K}^{-1}, \tag{3.64}$$

and

$$\begin{aligned}(\Delta S)_{\mathrm{iso}} &= \Delta S_{\mathrm{syst}} + \Delta S_{\mathrm{surr}} \qquad (3.65)\\ &= (-35.42 + 36.82)\,\mathrm{J\,K}^{-1} = 1.40\,\mathrm{J\,K}^{-1}.\end{aligned}$$

The entropy change of such an isolated system is still bigger than zero, i.e. $(\Delta S)_{\mathrm{iso}} > 0$. Such an irreversible supercooling solidification does completely agree with the mathematical expression of the second law of thermodynamics.

Before ending the discussion in this section, it should be emphasized that classical thermodynamics really consists of two parts of equilibrium thermodynamics and nonequilibrium thermodynamics. The classical equilibrium thermodynamics can be used for any macroscopic equilibrium systems without any other assumption or premise, so the application of equilibrium thermodynamics is well developed in classical thermodynamics. However, the application of nonequilibrium thermodynamics is severely restricted in classical thermodynamics. That is the limitation of classical thermodynamics as discussed in Chapter 2. Irreversible processes, which can be treated in classical thermodynamics, must accord with the assumption of initial and final equilibrium states, as shown in Fig. 3.1. For meeting the requirement of Clausius assumption of initial and final equilibrium states, the use of clapboard and drawing the clapboard away have to be introduced to ensure that both the initial and final states are in equilibrium. The next step is to design a reversible process by which the initial state and the final state could be connected. In fact, that also meets the requirement of Clausius cyclical-process assumption. According to these assumptions, beside Fig. 3.4, Fig. 3.5 and Fig. 3.7 there are also chemical reactions, as shown in Fig. 3.8, and so on.

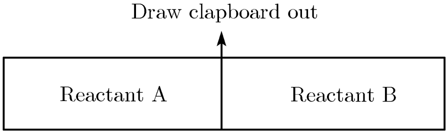


Fig. 3.8 An irreversible process of chemical reactions

To find a reversible process to realize a cyclical process for such kinds of chemical reactions it is necessary to use the equilibrium box designed by Jacobus Henricus van't Hoff (1852–1911), as shown in Fig. 3.9. In Fig. 3.9 there is a semipermeable membrane for each chemical component. Each chemical component may add into or take out at its equilibrium pressure. Such an equilibrium box can be used for gas mixing as well. The "isotherm of chemical reactions" in thermodynamics, which will be discussed in the following part of this chapter, was also contributed by van't Hoff. His contributions, such as the relationship between optical activity and the presence of an asymmetrical carbon atom and so on, were not only in thermodynamics. In 1901, the very first Nobel Prize in Chemistry was awarded to Jacobus H. van't Hoff (Fig. 3.10)[3] "in recognition of the extraordinary services he has rendered by the discovery of the laws of chemical dynamics and osmotic pressure in solutions".

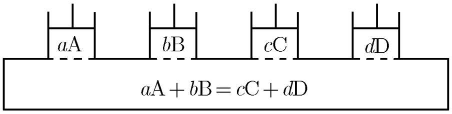


Fig. 3.9 Equilibrium box designed by Van't Hoff for realizing a reversible process of chemical reactions

Fig. 3.10 Jacobus Henricus van't Hoff (1852–1911)

There is a short story as follows. At the end of 1946 at the Brussels IUPAP (International Union of Pure and Applied Physics) meeting, after Prigogine had made a presentation on the thermodynamics of irreversible processes a specialist of great repute said to him: "I am surprised that you give more attention to irreversible phenomena, which are essentially transitory, than to the final result of their evolution, equilibrium." [4] [14] This was a typical viewpoint of classical thermodynamics from which one may clearly know the limitation of classical thermodynamics. At the same time we know that classical thermodynamics is not suitable for most of practical nonequilibrium states (including stationary nonequilibrium states), which are impersonally existent in a large amount everywhere and not to be evanescence. From the discussions in this section we can clearly know that any quan-

---

[3] http://nobelprize.org/nobel_prizes/chemistry/laureates/1901/hoffbio.html. Cited 20 Apr 2008

[4] http://nobelprize.org/nobel_prizes/chemistry/laureates/1901/hoffbio.html. Cited 20 Apr 2008

titative calculation must be on the basis of equilibrium systems or reversible processes in classical thermodynamics. That is a severe limitation of classical thermodynamics.

3.7 Relationship between Gibbs Free Energy and T or p

In nature and also in the productive or research work of human beings there are a lot of isothermal isobaric reactions or processes. More detailed discussions are needed on the properties of the Gibbs free energy, such as the relationship between the function of G and T or p. For a pure substance, 1 mole is usually taken as a unit. The relationship between its G value and temperature can be derived as follows. According to the basic thermodynamic formula, we know

$$\left(\frac{\partial G}{\partial T}\right)_p = -S. \tag{3.66}$$

From the definition of G, $G = H - TS$, we get

$$\left(\frac{\partial G}{\partial T}\right)_p = \frac{G-H}{T} \tag{3.67}$$

or

$$\frac{1}{T}\left(\frac{\partial G}{\partial T}\right)_p - \frac{G}{T^2} = -\frac{H}{T^2}, \tag{3.68}$$

that is,

$$\left[\frac{\partial\left(\frac{G}{T}\right)}{\partial T}\right]_p = -\frac{H}{T^2}. \tag{3.69}$$

After transposition and integration,

$$\int \mathrm{d}\left(\frac{G}{T}\right) = \int\left(-\frac{H}{T^2}\right)\mathrm{d}T, \tag{3.70}$$

$$\frac{G}{T} = -\int \frac{H}{T^2}\mathrm{d}T + I. \tag{3.71}$$

Here, I is the integration constant.

If the isobaric heat capacity, C_p, can be expressed as a function of temperature, such as $C_p = A + BT + CT^{-2} + DT^2$, then

$$H = H_0 + \int C_p \mathrm{d}T = H_0 + AT + \frac{1}{2}BT^2 - CT^{-1} + \frac{1}{3}DT^3. \tag{3.72}$$

Here, H_0 is the integration constant. By substitution of G/T in Eq. (3.71), after transposition and integration, we get

$$\frac{G}{T} = \frac{H_0}{T} - A\ln T - \frac{1}{2}BT - \frac{1}{2}CT^{-2} - \frac{1}{6}DT^2 + I. \tag{3.73}$$

That is,

$$G = H_0 - AT\ln T - \frac{1}{2}BT^2 - \frac{1}{2}CT^{-1} - \frac{1}{6}DT^3 + IT. \tag{3.74}$$

From thermodynamic data books it is possible to obtain the relation of C_p–T for a given substance, and also the standard values of H^0 and G^0 (or S^0) at the standard pressure and 298.15 K. From these values and the above-mentioned formulae, I and H^0 can firstly be obtained, and then all values of H^0, G^0 and S^0 at different temperature T and at standard pressure can be calculated in the valid range of the C_p–T relation. The superscript (^{0}) here represents only the condition of standard pressure and does not denote the temperature. In this book the enthalpy H^0 of any stable pure substance at 101.325 kPa and 298.15 K is relatively taken as zero.

According to the basic formula $\mathrm{d}G = S\mathrm{d}T + V\mathrm{d}p$, the relationship between the function G and the pressure for pure substances can be derived from

$$\left(\frac{\partial G}{\partial p}\right)_T = V. \tag{3.75}$$

After transposition and integration, we get

$$G(p_2,T) - G(p_1,T) = \int_{p_1}^{p_2} V\mathrm{d}p. \tag{3.76}$$

If p_1 is the standard pressure, $p_1 = p^0$, then the G value at pressure p can be obtained from the following equation.

$$G(p,T) = G^0(p^0,T) + \int_{p^0}^{p} V\mathrm{d}p. \tag{3.77}$$

If the substance is an ideal gas and the amount equals 1 mole, then

$$G_\mathrm{m}(p,T) = G^0_\mathrm{m}(p^0,T) + RT\ln\left(\frac{p}{p^0}\right). \tag{3.78}$$

If the $G^0(p^0,T)$ value at T and at standard pressure is known, then the $G(p,T)$ value at other pressures can be calculated. The subscript “m” here represents 1 mole of substance.

If 1 mole of a solid pure substance, e.g. graphite, at 298.15 K undergoes a pressure change from an initial pressure of 1 atmosphere to 100 atmospheres; then the change of G value can be calculated by the following steps.

$$V_\mathrm{m} = \frac{M}{\rho} = \frac{12.011\mathrm{gmol}^{-1}}{2.260\mathrm{gcm}^{-3}}. \tag{3.79}$$

Here M is the atomic weight of carbon. It is assumed that the density ρ is a constant in the relevant pressure range. Therefore,

$$\begin{aligned} \Delta G_\mathrm{m} &= (12.011/2.260) \times (100-1) \times 101.325\mathrm{kPacm^3mol^{-1}} \\ &= 53\mathrm{Jmol^{-1}}. \end{aligned} \tag{3.80}$$

The G value for 1 mole of a pure substance is usually of the order of kJ mol^{-1}, so such a small value of 53 J mol^{-1} is negligible. Usually, chemical reactions run at about 1 atmosphere, so the influence of a pressure change on the G value of a solid substance is negligible. However, in high-pressure physics the pressure change may be of the order of GPa. In such conditions, the influence of the pressure on the solid should be under consideration. That will be discussed in the following chapters.

3.8 Relationship between Chemical Potential and T or p

For multicomponent systems, the composition of the system, i.e. the number of moles of each component, may change. That may be the result of the influence of matter flow in an open system, or the results of the reaction or phase transformation in a closed system. During 1875–1878, Gibbs published two parts of his famous paper, "*On the Equilibrium of Heterogeneous Substances*", in which he introduced the concept of chemical potential for multicomponent systems. This paper is generally considered as his most important contribution to physical sciences. As mentioned in Chapter 2, before his article was translated into German by Ostwald in 1892 and into French by la Chatlier in 1899, Gibbs work did not generally attract other scientists in Europe. Many contents of equilibrium thermodynamics today can be traced back to the important literature of Gibbs. (p. 123) [2,15]

The properties of a multicomponent system can be represented by any two independent variables of the thermodynamic functions p, T, V, U, S, $\cdots$, and the amount of each component. For instance, for the Gibbs free energy of system, T, p, n_1, n_2, $\cdots$, n_k can be selected as independent variables, i.e. $G = G(T, p, n_1, n_2, \cdots, n_k)$. If there are k components in the system and the numbers of moles of each component are n_1, n_2, $\cdots$, n_k, respectively, then the total differential of G can be written as

$$\begin{aligned} \mathrm{d}G &= \left(\frac{\partial G}{\partial T}\right)_{p,n_j} \mathrm{d}T + \left(\frac{\partial G}{\partial p}\right)_{T,n_j} \mathrm{d}p + \sum_{i=1}^{k} \left(\frac{\partial G}{\partial n_i}\right)_{T,p,n_j} \mathrm{d}n_i, \qquad (3.81) \\ &= -S\mathrm{d}T + V\mathrm{d}p + \sum_{i=1}^{k} \left(\frac{\partial G}{\partial n_i}\right)_{T,p,n_j} \mathrm{d}n_i, \end{aligned}$$

where n_j represents the number of moles of all other components except n_i. The definition for the last term can be given as

$$\mu_i = \left(\frac{\partial G}{\partial n_i}\right)_{T,p,n_j}. \qquad (3.82)$$

μ_i is called the chemical potential of component i. When the temperature T, pressure p and all other components n_j keep constant, the increase of the Gibbs free energy of a huge system due to the addition of 1 mole of substance i is called the chemical potential of substance μ_i. The system should be so huge that concentrations of all components remain unchanged during the addition.

It should be noted that the appropriate independent parameters should be selected for different thermodynamic functions. The following four partial differentials can be called chemical potentials, if the respective subscripts remain constant. That is

$$\mu_i = \left(\frac{\partial U}{\partial n_i}\right)_{S,V,n_j} = \left(\frac{\partial H}{\partial n_i}\right)_{S,p,n_j}$$
$$= \left(\frac{\partial F}{\partial n_i}\right)_{T,V,n_j} = \left(\frac{\partial G}{\partial n_i}\right)_{T,p,n_j}. \tag{3.83}$$

It should also be carefully noted that not all partial differentials could be called chemical potentials. Therefore, for multicomponent and composition variable systems, the four basic formulae of thermodynamics Eq. (3.16)–Eq. (3.19) should be rewritten as

$$\mathrm{d}U = T\mathrm{d}S - p\mathrm{d}V + \sum_i \mu_i \mathrm{d}n_i, \tag{3.84}$$

$$\mathrm{d}H = T\mathrm{d}S + V\mathrm{d}p + \sum_i \mu_i \mathrm{d}n_i, \tag{3.85}$$

$$\mathrm{d}F = -S\mathrm{d}T - p\mathrm{d}V + \sum_i \mu_i \mathrm{d}n_i, \tag{3.86}$$

and

$$\mathrm{d}G = -S\mathrm{d}T + V\mathrm{d}p + \sum_i \mu_i \mathrm{d}n_i. \tag{3.87}$$

The most general condition for running reactions or processes is at constant temperature and pressure, so the most general form of $\mu_i = \left(\frac{\partial G}{\partial n_i}\right)_{T,p,n_j}$ is often used. If there is no special mention, the chemical potential usually represents the chemical potential expression of the Gibbs free energy. The subscripts T, p and n_j of $\left(\frac{\partial G}{\partial n_i}\right)_{T,p,n_j}$ mean that the reaction or process is running at constant temperature, pressure and amounts of other components. It should be noted that only $\left(\frac{\partial G}{\partial n_i}\right)_{T,p,n_j}$ can also be called the partial molar differential of the Gibbs free energy, but the other expressions of the chemical potential cannot be called the partial molar differentials, because the other expressions are not suitable at constant temperature and pressure.

It is very convenient to solve problems concerning multiphase equilibrium by the introduction of the concept of chemical potential. For instance, there are phase α and phase β in a system. If a small amount of substance i transforms from phase α to phase β then the change of the total Gibbs free energy of the system is

$$\mathrm{d}G = \mathrm{d}G^{\alpha} + \mathrm{d}G^{\beta} = \mu_i^{\alpha}\mathrm{d}n_i^{\alpha} + \mu_i^{\beta}\mathrm{d}n_i^{\beta}. \tag{3.88}$$

Here, $\mathrm{d}n_i^{\alpha} = -\mathrm{d}n_i^{\beta}$. If phase equilibrium has already been reached in the system, then $\mathrm{d}G = 0$ and $\mu_i^{\alpha}\mathrm{d}n_i^{\alpha} = \mu_i^{\beta}\mathrm{d}n_i^{\alpha}$, i.e. $\mu_i^{\alpha} = \mu_i^{\beta}$. The chemical potentials of substance i in the two phases must be equal in equilibrium. The conclusion can be extensively used for any cases of multiphase systems. A similar conclusion can also be used for reaction systems. If a chemical reaction at constant temperature and pressure is

$$a\mathrm{A} + b\mathrm{B} + \cdots \rightarrow f\mathrm{F} + g\mathrm{G} + \cdots. \tag{3.89}$$

It can be shown from Eq. (3.87) that

$$\mathrm{d}G = \sum_i \mu_i \mathrm{d}n_i. \tag{3.90}$$

All amounts of individual substances change simultaneously and relatively in a chemical reaction. If the concept of the extent of reaction ξ is introduced, then

$$\mathrm{d}\xi = \frac{\mathrm{d}n_\mathrm{F}}{f} = \frac{\mathrm{d}n_\mathrm{G}}{g} = \cdots = -\frac{\mathrm{d}n_\mathrm{A}}{a} = -\frac{\mathrm{d}n_\mathrm{B}}{b}, \tag{3.91}$$

$$\mathrm{d}\xi = \frac{\mathrm{d}n_i}{\nu_i} \mathrm{or} n_i = \nu_i \mathrm{d}\xi. \tag{3.92}$$

Therefore,

$$\mathrm{d}G = \sum_i \nu_i \mu_i \mathrm{d}\xi. \tag{3.93}$$

Here, ν_i represents the stoichiometric coefficient of component i. Suppose that the system is huge enough, only 1 mole of the reaction equation has taken place [i.e. a moles of A react with b moles of B (and other reactants) in Eq. (3.89)] and concentrations of all components remain constant. Then

$$\Delta G_\mathrm{m} = \sum_i \nu_i \mu_i. \tag{3.94}$$

ΔG_m represents the change of the Gibbs free energy of the system, after 1 mole of the reaction equation has taken place. In chemical equilibrium, the sum of the chemical potentials of all productions equals the sum of the chemical potentials of all reactants. This can be written as

$$\Delta G_m = \sum_i \nu_i \mu_i = 0. \tag{3.95}$$

The relationship between chemical potential and temperature can be deduced as follows:

$$\begin{aligned} \left(\frac{\partial \mu_i}{\partial T}\right)_{p,n_i,n_j} &= \left[\frac{\partial}{\partial T}\left(\frac{\partial G}{\partial n_i}\right)_{T,p,n_j}\right]_{p,n_i,n_j} \\ &= \left[\frac{\partial}{\partial n_i}\left(\frac{\partial G}{\partial T}\right)_{p,n_i,n_j}\right]_{T,p,n_j} \\ &= \left[\frac{\partial}{\partial n_i}(-S)\right]_{T,p,n_j} = -\left(S_\mathrm{m}\right)_i, \end{aligned} \tag{3.96}$$

where $(S_\mathrm{m})_i$ represents the partial differential molar entropy of substance i. By partial differentiation of $G = H - TS$ with respect to n_i at constant temperature and pressure,

$$\left(\frac{\partial G}{\partial n_i}\right)_{T,p,n_j} = \left(\frac{\partial H}{\partial n_i}\right)_{T,p,n_j} - T\left(\frac{\partial S}{\partial n_i}\right)_{T,p,n_j} \tag{3.97}$$

or

$$\mu_i = (H_\mathrm{m})_i - T\,(S_\mathrm{m})_i\,. \tag{3.98}$$

Note that $\left(\frac{\partial H}{\partial n_i}\right)_{T,p,n_j} = (H_\mathrm{m})_i$ and $\left(\frac{\partial S}{\partial n_i}\right)_{T,p,n_j} = (S_\mathrm{m})_i$ are partial molar quantities at constant temperature and pressure, but not chemical potentials. Only $\left(\frac{\partial G}{\partial n_i}\right)_{T,p,n_j} = \mu_i$ is the chemical potential and partial molar quantity. Correspondingly,

$$\begin{aligned}\left[\frac{\partial\left(\frac{\mu_i}{T}\right)}{\partial T}\right] &= \frac{T\left(\frac{\partial \mu_i}{\partial T}\right)_{p,n_i,n_j} - \mu_i}{T^2} \\ &= \frac{-\left[T\,(S_\mathrm{m})_i + \mu_i\right]}{T^2} = -\frac{(H_\mathrm{m})_i}{T}.\end{aligned} \tag{3.99}$$

The relation between chemical potential and pressure is

$$\begin{aligned}\left(\frac{\partial \mu}{\partial p}\right)_{T,n_i,n_j} &= \left[\frac{\partial}{\partial p}\left(\frac{\partial G}{\partial n_i}\right)_{T,p,n_j}\right]_{T,n_i,n_j} \\ &= \left[\frac{\partial}{\partial n_i}\left(\frac{\partial G}{\partial p}\right)_{T,n_i,n_j}\right]_{T,p,n_j} = \left(\frac{\partial V}{\partial n_i}\right)_{T,p,n_j} = (V_\mathrm{m})_i\,.\end{aligned} \tag{3.100}$$

For an ideal gas,

$$\mu\,(T,p) = \mu^0\,(T) + RT\ln\left(\frac{p}{p^0}\right) \tag{3.101}$$

or for substance i in an ideal gaseous mixture,

$$\mu_i\,(T,p) = \mu_i^0\,(T) + RT\ln\left(\frac{p_i}{p^0}\right). \tag{3.102}$$

The corresponding pressure of $\mu^0\,(T)$ or $\mu_i^0\,(T)$ is the standard pressure, so both $\mu^0\,(T)$ and $\mu_i^0\,(T)$ are only a function of temperature.

To sum up, the starting points for getting the relation between chemical potential and temperature or pressure are

$$\left(\frac{\partial \mu_i}{\partial T}\right)_{p,n_i,n_j} = -\,(S_\mathrm{m})_i$$

and

$$\left(\frac{\partial \mu_i}{\partial p}\right)_{T,n_i,n_j} = (V_\mathrm{m})_i$$

respectively.

In comparison with the corresponding equations for the pure substance, it can be found that the thermodynamic equations for both the multicomponent system and the system of a pure substance are similar in form. The distinction between them is that the molar quantities of pure substances have been replaced by partial molar differentials. In other words, for a pure substance its partial molar differentials equal its molar quantities. That is to say, the chemical potential of a pure substance is just the same as its 1 mole Gibbs free energy G_{m}.

3.9 Gibbs Free Energy Changes of Chemical Reactions

The last two sections have discussed the relationship between the chemical potential of a pure substance or one substance in a multicomponent system and temperature or pressure. If a chemical reaction takes place inside a system at constant temperature and pressure, then the change of the Gibbs free energy during the reaction should be taken as a thermodynamic criterion for the spontaneous direction of the reaction. Therefore, the change of the Gibbs free energy after one mole amount of the reaction has taken place is an important parameter, the negative value of which was defined by Th. De Donder as the chemical affinity A. The chemical affinity is more usually called the driving force of a reaction, and denoted by D.

$$A = D = -\Delta G_{\mathrm{m}} = -\sum_i \nu_i \mu_i. \tag{3.103}$$

If $A > 0$, the spontaneous direction is forward to the right side. If $A < 0$, the spontaneous direction is backward to the left side. If $A = 0$, the reaction may be in equilibrium. Obviously, the relationship ΔG or ΔG_{m} of the reaction with temperature or pressure should be clearly understood. Similarly to the last two sections, from the starting point of $\left(\frac{\partial \Delta G}{\partial T}\right)_p = -\Delta S$ and $\Delta G = \Delta H - T\Delta S$, the relationship of the Gibbs free energy change of reaction with temperature similar to that of a pure substance can be deduced. That is, the following equations can be obtained. ΔC_p is the difference of isobaric heat capacities between products and reactants.

$$\Delta C_p = \Delta A + \Delta BT + \Delta CT^{-2} + \Delta DT^2, \tag{3.104}$$

$$\begin{aligned} \Delta H &= \Delta H_0 + \int \Delta C_p dT \\ &= \Delta H_0 + \Delta AT + \frac{1}{2}\Delta BT^2 - \Delta CT^{-1} + \frac{1}{3}\Delta DT^3 \end{aligned} \tag{3.105}$$

and

$$\begin{aligned} \Delta G = \Delta H_0 - \Delta AT \ln T - \frac{1}{2}\Delta BT^2 \\ -\frac{1}{2}\Delta CT^{-1} - \frac{1}{6}\Delta DT^3 + \Delta IT. \end{aligned} \tag{3.106}$$

If the values ΔH, ΔG at 298.15K and the relation ΔC_p–T are known, the two integration constants of ΔH_0 and ΔI can be obtained from the equations listed above, and then all the values of ΔH, ΔG and ΔS (based on $\Delta H = \Delta G - T\Delta S$) at different temperatures in the valid range of ΔC_p–T can be obtained by substituting T. If all reactants and products in the reaction are individually at the standard pressure (101.325 kPa), the symbols ΔH^0 and ΔG^0 will be used instead ΔH and ΔG.

Consider again the chemical reaction Eq. (3.89), i.e.

$$a\mathrm{A}+b\mathrm{B}+\cdots \rightarrow f\mathrm{F}+g\mathrm{G}+\cdots.$$

Suppose that all reactants and products are ideal gases. By substituting Eq. (3.102) in Eq. (3.94),

$$\Delta G_{\mathrm{m}}=\sum_i \nu_i\mu_i=\sum_i \nu_i\mu_i^0(T)+\sum_i \nu_i RT\ln\left(\frac{p_i}{p^0}\right). \tag{3.107}$$

If all reactants and products are at standard pressure, ΔG_{m}^0 should be used instead of ΔG_{m}.

$$\Delta G_{\mathrm{m}}^0=\sum_i \nu_i\mu_i^0(T), \tag{3.108}$$

$$\Delta G_{\mathrm{m}}=\Delta G_{\mathrm{m}}^0+RT\ln\frac{\left(\frac{p_{\mathrm{F}}}{p^0}\right)^f\left(\frac{p_{\mathrm{G}}}{p^0}\right)^g\cdots}{\left(\frac{p_{\mathrm{A}}}{p^0}\right)^a\left(\frac{p_{\mathrm{B}}}{p^0}\right)^b\cdots}, \tag{3.109}$$

where p_i represents the partial pressure of component i. When the system is in equilibrium, $\Delta G_{\mathrm{m}}=0$, so

$$\Delta G_{\mathrm{m}}^0=-RT\ln\frac{\left(\frac{p_{\mathrm{F}}^e}{p^0}\right)^f\left(\frac{p_{\mathrm{G}}^e}{p^0}\right)^g\cdots}{\left(\frac{p_{\mathrm{A}}^e}{p^0}\right)^a\left(\frac{p_{\mathrm{B}}^e}{p^0}\right)^b\cdots}. \tag{3.110}$$

K_p^{o} is defined as a number without dimensions, and is called the thermodynamic equilibrium constant:

$$K_p^{\mathrm{o}}=\frac{\left(\frac{p_{\mathrm{F}}^e}{p^0}\right)^f\left(\frac{p_{\mathrm{G}}^e}{p^0}\right)^g\cdots}{\left(\frac{p_{\mathrm{A}}^e}{p^0}\right)^a\left(\frac{p_{\mathrm{B}}^e}{p^0}\right)^b\cdots}, \tag{3.111}$$

so

$$\Delta G_m^0=-RT\ln K_p^{\mathrm{o}} or \Delta G^0=-nRT\ln K_p^{\mathrm{o}}. \tag{3.112}$$

Here, p^0 represents the standard pressure (100.325 kPa), and p_i^e represents the equilibrium partial pressure of component i. It is defined so that

$$Q_p=\frac{\left(\frac{p_{\mathrm{F}}}{p^0}\right)^f\left(\frac{p_{\mathrm{G}}}{p^0}\right)^g\cdots}{\left(\frac{p_{\mathrm{A}}}{p^0}\right)^a\left(\frac{p_{\mathrm{B}}}{p^0}\right)^b\cdots}. \tag{3.113}$$

Therefore,

$$\Delta G_{\mathrm{m}}=-RT\ln K_p^{\mathrm{o}}+RT\ln Q_p, \tag{3.114}$$

$$\Delta G=-nRT\ln K_p^{\mathrm{o}}+nRT\ln Q_p, \tag{3.115}$$

$$\Delta G_{\mathrm{m}}=\Delta G_{\mathrm{m}}^0+RT\ln Q_p \tag{3.116}$$

or

$$\Delta G = \Delta G^0 + nRT \ln Q_p. \tag{3.117}$$

This is the complete relationship between the equilibrium constant and ΔG of a practical chemical reaction, which was first deduced by J.H. van't Hoff in 1886 and called the "reaction isotherm" or the "isotherm of chemical reactions".

It should be noted that

1. The thermodynamic equilibrium constant K_p° is a number without dimensions, which is different from the equilibrium constants K_p in many textbooks of chemistry. The ordinary equilibrium constants K_p are usually defined as

$$K_p = \frac{(p_{\mathrm{F}}^e)^f\,(p_{\mathrm{G}}^e)^g \cdots}{(p_{\mathrm{A}}^e)^a\,(p_{\mathrm{B}}^e)^b \cdots}, \tag{3.118}$$

   where K_p is a number with dimensions. The value of K_p changes with the unit used for pressure, while the thermodynamic equilibrium constant K_p° is independent of the pressure unit.
2. The superscript (^{0}) of the thermodynamic equilibrium constant K_p° does not represent the standard pressure here, because from Eq. (3.108) and Eq. (3.112) K_p° is only a temperature relative function and independent of pressure.
3. p_{A}, p_{B}, $\cdots$, p_{F}, p_{G}, $\cdots$ are practical partial pressures of the reactants and products. ΔG^0 means the Gibbs free energy change of the reaction when all pressures of reactants and products are individually at standard pressure. The term ΔG^0 does not relate to the practical pressures of reactants and products.

It should be emphasized that ΔG^0 should never be used as a strict criterion of any practical reaction direction instead of ΔG. The key difference between ΔG^0 and ΔG is their dependence on practical partial pressures (or practical concentrations). The former (the value of ΔG^0) is independent of practical partial pressures and does not connect with the degree of advancement of the reaction, while the latter (the value of ΔG) is closely dependent on the practical partial pressures and changes with the degree of advancement of the reaction. When ΔG equals zero and reaches its minimum value, the system is in equilibrium.

From the equations $\Delta G_{\mathrm{m}}^0 = -RT \ln K_p^{\circ}$ and $\left[\frac{\mathrm{d}\left(\frac{\Delta G_{\mathrm{m}}^0}{T}\right)}{\mathrm{d}T}\right] = -\frac{\Delta H_{\mathrm{m}}^0}{T^2}$, it is easy to obtain

$$\frac{\mathrm{d}\ln K_p^{\circ}}{\mathrm{d}T} = \frac{\Delta H_{\mathrm{m}}^0}{RT^2} \tag{3.119}$$

and

$$\ln K_p^{\circ} = -\frac{\Delta H_0}{RT} + \frac{\Delta A}{R}\ln T + \frac{\Delta B}{2R}T + \frac{\Delta C}{2R}T^{-2} + \frac{\Delta D}{6R}T^2 + \frac{\Delta I}{R}. \tag{3.120}$$

If ΔH_{m}^0 and K_p° at a definite temperature, and the relationship C_p–T in a certain range are known, then from these equations the two integration constants ΔH_0 and ΔI can be obtained, and then the value of K_p° at different temperature can be calculated.

For a reaction system consisting of ideal gases, in the derivation of the reaction equilibrium constant via chemical potentials one finds

$$RT\ln K_p^{\mathrm{o}} = \sum_i \nu_i \mu_i^0 (T) = \Delta G_{\mathrm{m}}^0. \tag{3.121}$$

$\mu_i^0 (T)$ is only a function of temperature, so

$$\left[\frac{\partial \ln K_p^{\mathrm{o}}}{\partial p}\right]_T = 0. \tag{3.122}$$

For condensed systems, the starting point for the derivation of the relationship between ΔG_{m} and pressure is the following equation:

$$\left[\frac{\partial (\Delta G_{\mathrm{m}})}{\partial p}\right]_T = \Delta V_{\mathrm{m}}. \tag{3.123}$$

After transposition and integration from the standard pressure to an arbitrary pressure, we may obtain

$$\Delta G_{\mathrm{m}} = \Delta G_{\mathrm{m}}^0 + \int_{p^0}^{p} \Delta V_{\mathrm{m}} \mathrm{d}p. \tag{3.124}$$

Suppose that in condensed systems the densities of the reactants and products does not change with pressure, so

$$\Delta G_{\mathrm{m}} = \Delta G_{\mathrm{m}}^0 + \Delta V_{\mathrm{m}} \left(p - p^0\right). \tag{3.125}$$

Based on this relation, the density data of diamond and graphite, and the value of ΔG_{m}^0, it can be estimated how high a pressure is needed to achieve the transformation from graphite to diamond. That will be discussed in detail later.

At the end of this chapter, it should be emphasized that "what is classical thermodynamics" should be the most fundamental concept, but due to the historical reasons the answer will more clearly be found in the next chapter, or even after reading the whole book.

References

1. Clausius R. Mechanical Theory of Heat [M]. London: John van Voorst, 1867.
2. Kondepudi D, Prigogine I. Modern Thermodynamics — From Heat Engines to Dissipative Structures [M]. New York: John Wiley & Sons, 1998. http://www.wiley.com/. Cited on 20 Apr 2008.
3. Fu X-C, Shen W-X, Yao T-Y. Physical Chemistry [M]. 4th edn. Beijing: High Education Press, 1990.
4. Massieu MF. Compt Rend [J], 1869, 69: 858, 1057.
5. Wang J-T. Modern Thermodynamics — and a whole view of thermodynamics [M]. Shanghai: Fudan Univ Press, 2005, 223 pages.
6. Wang J-T. Nonequilibrium Nondissipative Thermodynamics — With Application to Low-Pressure Diamond Synthesis [M]. Heidelberg: Springer, 2002.
7. Wang J-T. Physics (Beijing) [J], 2003, 32(1): 9-15.
8. Wang J-T. Phase Diagrams of Stationary Nonequilibrium States — Thermodynamics for Low Pressure Diamond Growth from the Vapor Phase [M]. Beijing: Science Press, 2000, 212 pages.

9. Wang J-T, Zhang D W, Liu Z-J. Thermodynamic Coupling Model for Low Pres-sure Diamond Growth from the Vapor Phase [M]. Beijing: Science Press, 1998, reprinted in 2000, 175 pages.
10. Wang J-T. Modern Physics Lett B [J], 2002, 16(23&24): 885-888.
11. Hillert M. Phase Equilibria, Phase Diagrams and Phase Transformations Their Thermodynamic Basis [M]. Boston: Cambridge Univ Press, 1998.
12. Wang Z-X. Thermodynamics [M]. Beijing: High Education Press, 1955.
13. Su R-K. Statistic Physics [M]. 2nd edn. Beijing: High Education Press, 2004.
14. Shen X-F, Hu G, Jiang L. Theory of Dissipative Structures [M]. Shanghai: Shanghai People's Publisher, 1987.
15. Gibbs J W. On the Equilibrium of Heterogeneous Substances. Trans Conn Acad Sci [J], 1875, III: 108-248, 1878: 343-524; Gibbs J W. On the equilibrium of heterogeneous substances, 1876. In Collected Works, volume 1 [M]. London: Longmans Green and Co, 1928.

Chapter 4
Fundamentals of Modern Thermodynamics

Abstract Modern thermodynamics is a part of thermodynamics for complex systems, or more exactly, modern thermodynamics is a part of thermodynamics for researches on coupling systems, in which both nonspontaneous and spontaneous processes exist simultaneously. Because of the complicities of the simultaneous existence of nonspontaneous and spontaneous processes, modern thermodynamics had been blocked at its developing stage nearly through the whole of the 20th century. The term "modern thermodynamics" was not generally accepted nearly through the whole of the 20th century, or only meant the status of thermodynamics at that time. The interaction between simultaneous nonspontaneous and spontaneous processes in the same system had been called "compensation" by Clausius in 1865, but is now usually called thermodynamic coupling, which is the core of modern thermodynamics.

4.1 Introduction

In 1931, the proposal of Onsager reciprocal relations was out of traditional thermodynamics in the 19th century. The term "classical thermodynamics" was likely to have appeared in the literature around the 1930s, so that time or even the whole of the 20th century may be regarded as the exploration stage of modern thermodynamics. Onsager and Prigogine won their Nobel Prizes for their contributions of reciprocal relations and dissipative structures in 1968 and 1977, respectively. These were important achievements at that time. However, as said by themselves, during the deduction of Onsager reciprocal relations out-of-thermodynamics "assumption of microscopic reversibility" (or assumption of "detailed balancing") had to be introduced, and Prigogine dissipative structures were established on the basis of an out-of-thermodynamics nonlinear positive feedback kinetic model, called "Brusselator". Therefore, Onsager reciprocal relations are approximate, while Prigogine dissipative structures are qualitative statements in written descriptions.

As pointed out in the brief introduction of the book, "*Modern Thermodynamics — from heat engines to dissipative structures*" by D. Kondepudi and I. Prigogine (1998) [1] in the website of Wiley Publisher during 2000 – 2002.[1]

> Thermodynamics is a core part of science. Most books present thermodynamics very much the way it was presented in the 19th century.

After 2004, such a brief introduction had been changed a little as follows.[2]

---

[1] http://www.wiley.com/WileyCDA/WileyTitle/productCd0471973947.html. Cited 2000–2002

[2] http://www.wiley.com/WileyCDA/WileyTitle/productCd0471973947.html. Cited 2004–2008

> Thermodynamics is a core part of most science and engineering curricula. However, most texts that are currently available to students still treat thermodynamics very much as it was presented in the 19th century, generally for historical rather than pedagogical reasons.

That is a severe comment on the real situation of thermodynamics. It means that most scientists had received such an undergraduate and/or graduate education in classical thermodynamics (perhaps even only classical equilibrium thermodynamics) in the last 150 years or so. Maybe it is a double-edged sword, as it also means a severe situation for modern thermodynamics that after about a century of exploration the term “modern thermodynamics” had not been well accepted. This may be due to some basic problems or discrepancies inside “modern thermodynamics”. The situation from both sides had severely hindered the application of thermodynamics in modern science and technology, and also the development of modern thermodynamics itself.

In this chapter a basic and complete modern classification of thermodynamics will be proposed by the author on the basis of our 20-year research on thermodynamics for activated low-pressure diamond growth from the vapor phase, or simply say, on the basis of the extended Carnot theorem as written in the title of this book. On the other hand, this chapter will provide a theoretical basis of modern thermodynamics. The detailed applications of dissipative thermodynamics and nondissipative thermodynamics will be discussed in the following chapters.

Many fundamental concepts of modern thermodynamics are still coming from the fundamental concepts of classical thermodynamics, but the terms or concepts, such as thermodynamic coupling, coupling systems, uncoupling systems, dissipative thermodynamics, nondissipative thermodynamics, nonequilibrium phase diagrams, and nonequilibrium stationary phase diagrams, and so on, are the characteristic new terms in modern thermodynamics. Before the terms of “dissipative function” (see Section 4.4) and Prigogine dissipative structures, the term of “dissipative” or “dissipation” was rarely used directly. In modern thermodynamics these new terms and concepts are closely related to each other, so they will be elucidated in this chapter.

4.2 General Mathematical Expressions of Basic Laws

Classical thermodynamics had been concerned mainly with the study of isolated and closed systems. A striking achievement of developments in the 20th century was withdrawing this limitation so as to generalize the methods of thermodynamics to open systems which are of great importance for biological thermodynamics as well as for many other fields such as meteorology, geology and so on. Modern thermodynamics is often necessary for studying systems in these fields, in which open systems are usually concerned. The name of closed systems is coming from the viewpoint of exchange of matter, but in modern thermodynamics the viewpoint of exchange of energy may often be more important, so sometimes closed systems in modern thermodynamics are also called open systems. For convenience to study different kinds systems, the mathematical expressions of fundamental thermodynamic laws had better be revised into their new forms. These revised forms of the mathematical expressions of fundamental thermodynamic laws can be called the modern (or general) mathematical expressions of fundamental thermodynamic laws, while their original forms can be called the classical mathematical expressions of fundamental thermodynamic laws.

The change of internal energy in the first law of thermodynamics can be split into two parts.[2, 3] One part is connected with the change from inside which is denoted by the symbol $\mathrm{d_i}U$. Another part is connected with the exchange of energy (for open systems the exchange of matter should be under consideration as well) between a system and its surroundings, which is denoted by the symbol $\mathrm{d_e}U$. Therefore, the total change of internal energy of the system can be written as

$$\mathrm{d}U = \mathrm{d_e}U + \mathrm{d_i}U. \tag{4.1}$$

According to the conservation of energy, energy inside system is impossible to be procreated and destroyed. Therefore, the first law of thermodynamics can be rewritten as

$$\mathrm{d_i}U = 0. \tag{4.2}$$

For closed systems,

$$\mathrm{d}U = \mathrm{d_e}U = \delta Q - \delta W. \tag{4.3}$$

For the second law of thermodynamics, entropy is the most important state function. Entropy of system is of an extensive property. If a system consists of several parts, the total entropy equals the sum of the entropies of each part. Similar to the modern expression of the first law, the entropy change of system $\mathrm{d}S$ can be split into two parts.[1~10] One is the entropy production denoted by $\mathrm{d_i}S$ due to the irreversible processes inside the system. Another is the entropy flow denoted by $\mathrm{d_e}S$ due to interactions with the surroundings, such as exchanges of energy and/or matter. Therefore, the total entropy change of the system $\mathrm{d}S$ equals

$$\mathrm{d}S = \mathrm{d_e}S + \mathrm{d_i}S. \tag{4.4}$$

Entropy flow of a system $\mathrm{d_e}S$ may be positive or negative, but entropy production of any system $\mathrm{d_i}S$ is never negative; that is the positive entropy production principle:

$$\mathrm{d_i}S \geqslant 0. \tag{4.5}$$

For simple (uncoupling) systems, if systems with only reversible processes inside, then $\mathrm{d_i}S = 0$; while if systems with only spontaneous irreversible processes inside, then $\mathrm{d_i}S > 0$. For an isolated system, $\mathrm{d_e}S = 0$, so that $(\mathrm{d}S)_{\mathrm{iso}} = \mathrm{d_i}S \geqslant 0$; that is to say, for any isolated system, the modern expression results in the same conclusion as the classical expression of the second law. However, the modern expression of the second law, $\mathrm{d_i}S \geqslant 0$, can be used not only for isolated systems, but also for closed systems and open systems. Therefore, Eq. (4.5) is a general mathematical expression of the second law of thermodynamics.

For closed and open systems, entropy flow $\mathrm{d_e}S$ may be positive, negative or zero. Nevertheless, such a situation of [$\mathrm{d_i}S < 0$, $\mathrm{d_e}S > 0$ & $\mathrm{d}S \geqslant 0$] is impossible, because $\mathrm{d_i}S < 0$ has already violated the general mathematical expression of the second law of thermodynamics. It does also mean that “absorption” of entropy in a system cannot be compensated by a big enough positive entropy flow of the system.

As mentioned above, the general mathematical expression of the second law, $\mathrm{d_i}S \geqslant 0$, is suitable for any macroscopic isolated, closed or open system (be careful, for nonequilibrium states the local equilibrium approximation has already been included; see the next Section 4.3 of this chapter), and is suitable for each macroscopic part of any system. Therefore, if an isolated system is split into two macroscopic parts I and II, then $d_iS^I \geqslant 0$ and $d_iS^{II} \geqslant 0$. Here, d_iS^I and d_iS^{II} represent the entropy productions of part I and part II, respectively. The physical situation such that

$$d_iS^I < 0, d_iS^{II} > 0 \& d_iS \geqslant 0 \tag{4.6}$$

is excluded. Here, $d_iS = (d_iS^I + d_iS^{II})$ represents the entropy production of the system. We can therefore say that the "absorption" of entropy in one part, compensated by a sufficient "production" of entropy in another part of the system, is prohibited. This conclusion implies that in every macroscopic region of the system the entropy production due to irreversible processes is positive. The term "macroscopic region" refers to any region containing a number of molecules sufficiently large for microscopic fluctuations to be negligible. The interaction between two irreversible processes is possible if these processes take place in the same region and in the same interval of time.

Another physical situation must be strictly distinguished. If there are two different processes taking place simultaneously in the same system (or in the same macroscopic region of a system), and the entropy productions due to process(es) 1 and process(es) 2 are denoted by d_iS_1 and d_iS_2, respectively, then it is possible to have

$$d_iS_1 < 0, d_iS_2 > 0 \& d_iS \geqslant 0. \tag{4.7}$$

Because only the sum of $d_iS_1 + d_iS_2$ represents the total entropy production d_iS of the macroscopic system, $d_iS = d_iS_1 + d_iS_2 \geqslant 0$ completely accords with the modern expression of the second law of thermodynamics. Such a situation is called thermodynamic coupling between irreversible processes, or thermodynamic coupling for short. If process 1 and/or process 2 is chemical reaction(s) then it can be called reaction coupling. The introduction of the concept of thermodynamic coupling or reaction coupling is one of the most important events in the development of modern thermodynamics, but the concept of thermodynamic coupling, especially the concept of reaction coupling in chemical reactions, had been bogged down in controversy for about 70 years. Thermodynamic coupling had not been well quantitatively demonstrated before the success of the thermodynamic coupling model for the activated low-pressure diamond synthesis. The emergence of the activated low-pressure diamond synthesis in the 1970s provided a new driving force for the development of thermodynamics on the basis of sufficient experimental data. The details of thermodynamic coupling will be discussed in Section 4.5 and other sections (and chapters) in this book. Be careful, the term "reaction" in this book may sometimes have an extended meaning, including physical process; while the term of "process" may have an extended meaning, including chemical reaction as well.

For a clearer understanding of the meanings of the entropy flow d_eS and the entropy production d_iS, the calculation of the entropy flow in a closed system may be used as a starting point for discussion. The entropy flow in a closed isochoric system can be quantitatively calculated based on the heat absorbed or given up by the system. The calculation must be on the basis of a reversible process; that is

$$\mathrm{d_e}S = \frac{\delta Q}{T}. \tag{4.8}$$

For reversible process, the system is always in equilibrium state, i.e. $\delta Q = \delta Q_\mathrm{R}$. For calculating nonequilibrium state T is temperature of heat reservoir which supplies δQ to the system.

For isothermal isobaric closed systems, according to the definition of Gibbs function, Eq. (3.10) $G = H - TS$, $\mathrm{d}G = \mathrm{d}H - T\mathrm{d}S$ can be obtained. That is entropy change of a system $\mathrm{d}S = \frac{\mathrm{d}H}{T} - \frac{\mathrm{d}G}{T}$. Here, $\mathrm{d}H$ is the heat absorbed by the system under isothermal isobaric conditions. In comparison with Eq. (4.4) $\mathrm{d}S = \mathrm{d_e}S + \mathrm{d_i}S$, $\frac{\mathrm{d}H}{T}$ corresponds to entropy flow $\mathrm{d_e}S$, while $-\frac{\mathrm{d}G}{T}$ corresponds to entropy production $\mathrm{d_i}S$ due to internal irreversible process(es) (including chemical reaction).

$$\mathrm{d_i}S = -\frac{\mathrm{d}G}{T}. \tag{4.9}$$

This equation is very useful for calculating the entropy production of reactions taking place at constant temperature and pressure.

Therefore, Eq. (4.7) can be written into its Gibbs function form, i.e.

$$(\mathrm{d}G_1)_{T,p} > 0, (\mathrm{d}G_2)_{T,p} < 0\& \ (\mathrm{d}G)_{T,p} \leqslant 0. \tag{4.10}$$

Here, $(\mathrm{d}G)_{T,p} = \left[(\mathrm{d}G_1)_{T,p} + (\mathrm{d}G_2)_{T,p}\right]$.

Of course, Eq. (4.7), Eq. (4.9) and Eq. (4.10) can be rewritten into their integral forms, i.e.

$$\Delta_\mathrm{i}S_1 < 0,\ \Delta_\mathrm{i}S_2 > 0\ \&\ \Delta_\mathrm{i}S \geqslant 0. \tag{4.11}$$

Here, $\Delta_\mathrm{i}S = (\Delta_\mathrm{i}S_1 + \chi\Delta_\mathrm{i}S_2)$, and reaction rate ratio of reaction 2 to reaction 1 is $\chi = r_2/r_1$. Customarily, $\Delta_\mathrm{i}S_1$ is the entropy production due to accomplishment of a chemical formula of reaction 1, so the rate of reaction 1 may be different from that of reaction 2.

$$\Delta_\mathrm{i}S = -\frac{\Delta G}{T}. \tag{4.12}$$

$$(\Delta G_1)_{T,p} > 0,\ (\Delta G_2)_{T,p} < 0\ \&\ (\Delta G)_{T,p} \leqslant 0. \tag{4.13}$$

Here, $(\Delta G)_{T,p} = \left[(\Delta G_1)_{T,p} + \chi\,(\Delta G_2)_{T,p}\right]$, and reaction rate ratio of reaction 2 to reaction 1 is $\chi = r_2/r_1$.

For open systems, the entropy flow should include the contribution due to the exchange of matter as well. That is the distinction between closed and open systems. If no irreversible process takes place inside the system, then $\mathrm{d_i}S = 0$. For systems with a single irreversible processes taking place a qualitative result $\mathrm{d_i}S > 0$ can be obtained. The detailed calculation of $\mathrm{d_e}S$ and $\mathrm{d_i}S$ will be further discussed later.

Obviously, thermodynamic coupling, as shown in Eq. (4.7), is not considered by classical thermodynamics. For simplification, the corresponding isothermal isobaric and integral forms will not be further discussed here. The mathematical expression of thermodynamic coupling, Eq. (4.7), may also be called the general mathematical expression of the second law of modern thermodynamics because it can generally be used for isothermal isobaric systems, nonisothermal nonisobaric systems, isolated systems, nonisolated systems, and so on. Equation (4.7) [$\mathrm{d_i}S_1 < 0$, $\mathrm{d_i}S_2 > 0$ & $\mathrm{d_i}S \geqslant 0$] is completely included in the general mathematical expression of the second law of the whole thermodynamics discipline, i.e. Eq. (4.5) [$\mathrm{d_i}S \geqslant 0$]. Therefore, from Eq. (3.7) [$(\mathrm{d}S_\mathrm{p})_\mathrm{iso} = (\mathrm{d}S)_\mathrm{iso} \geqslant 0$] for isolated systems or Eq. (3.43) [$(\mathrm{d}G_\mathrm{p})_{T,p} = (\mathrm{d}G)_{T,p} \leqslant 0$] for isothermal isobaric systems in Chapter 3, it can be known that the general mathematical expression of the second law of classical thermodynamics should be written as follows:

$$\mathrm{d_i}S_\mathrm{p} = \mathrm{d_i}S \geqslant 0. \tag{4.14}$$

Here, $\mathrm{d_i}S_\mathrm{p}$ and $\mathrm{d_i}S$ are the entropy production of spontaneous process and the entropy production of a simple uncoupling system, respectively. The general mathematical expression of the second law of classical thermodynamics, i.e. Eq. (4.14), has already included Eq. (3.7) for isolated systems in classical thermodynamics. For purposes of simplification, other integral forms will not be further discussed.

4.3 Local Equilibrium Approximation

Local equilibrium approximation is the most fundamental approximation for treatment of nonequilibrium states in thermodynamics. If state functions are regarded as continuous functions near equilibrium states, the local equilibrium approximation has already been included.

The huge achievement of classical thermodynamics is well known. A rather complete academic discipline of classical thermodynamics had been strictly deduced from and based only on a few laws, principles or theorems summarized from experiences, which are mainly mathematical expressions of the first law and the second law of thermodynamics. In particular, a lot of formulations have been quantitatively deduced in classical equilibrium thermodynamics for the reason that equilibrium thermodynamics belongs to the equality part of mathematical formulation of the second law of thermodynamics. The conclusions and predictions can be checked by different kinds of experiments or observations. However, classical thermodynamics has met a series of unusual difficulties during the treatment of practical questions concerning nonequilibrium states or irreversible processes. Perhaps the first difficulty was that the expression of the second law for irreversible processes is usually an inequality. It was difficult to use for quantitative deductions and derivations. Besides, more basic difficulties were whether these classical thermodynamic functions and formulations were still valid for systems in nonequilibrium states or for systems including irreversible processes. Perhaps the first problem was how to characterize the research system in a nonequilibrium state.

For example, suppose that there is a system with a unified temperature gradient in it. The left end of the system is connected with a heat reservoir of 400 K, while the right end is connected with a heat reservoir of 300 K. What is the temperature at the middle of the system? Maybe one can quickly make an intuitive answer: “It is 350 K.” In fact, this is a rather complex question concerned with the definition of temperature and temperature measurement. In thermodynamics, the temperature of a body is defined and measured by the zeroth law of thermodynamics. The zeroth law requires that heat equilibrium should be reached first between the thermometer and the body for measurement. In a nonuniform heat field the body (consisting of a large number of particles) itself is not yet in equilibrium. What is the temperature of a system in its nonequilibrium state? Of course, in practice the intuitive answer of 350 K may still be acceptable. However, in academic discussion, temperature, pressure, and all state functions or parameters in equilibrium thermodynamics should be defined again for nonequilibrium systems. The simplest way is to accept the local equilibrium approximation.[1~3,6~9]

The key points of the local equilibrium approximation are:

1. The system at its nonequilibrium state may be split into a lot of units with small volume. From a macroscopic point of view, each volume unit is so small that the property of any point inside the unit can be used without any obvious error as the property of the unit. From a microscopic point of view, there are still a lot of particles (such as molecules, atoms etc.) inside each volume unit, so that the statistical requirements can still be satisfied and the statistical properties (temperature, pressure, entropy and so on) can still be represented.
2. From the viewpoint of time, it is supposed that at time t these small volume units are isolated from their surroundings. At time $t+\delta t$ the original nonequilibrium state at time t has already been in equilibrium, so that the system at time $t+\delta t$ can be treated by equilibrium thermodynamics. If the time interval δt is still very small in comparison with the time scale of the macroscopic change of the whole system, the thermodynamic functions or parameters of each small volume unit at time t can be represented by the corresponding thermodynamic functions or parameters after equilibrium has been reached at time $t+\delta t$.
3. In addition, it should be supposed that the relationship among the obtained approximate thermodynamic functions or parameters is the same as the relationship among the corresponding functions or parameters in equilibrium thermodynamics.

After the local equilibrium approximation has been made, the system at its nonequilibrium state can be described and treated by the method similar to that in classical equilibrium thermodynamics. It is obvious that the local equilibrium approximation is not deduced from the basic laws of thermodynamics. Its suitable range is of a certain limitation. Overall, in such nonequilibrium systems the gradient of the thermodynamic functions or parameters should not be too big. For instance, for a dilute gas the change of thermodynamic functions or parameters in the range of the mean free path of molecules l should be much smaller than the mean value of the corresponding functions or parameters inside the range. For temperature and pressure, the requirements are

$$\frac{\Delta T}{T} \approx \frac{l|\nabla T|}{T} \ll 1 \tag{4.15}$$

and

$$\frac{\Delta p}{p} \approx \frac{l|\nabla p|}{p} \ll 1. \tag{4.16}$$

Here, ΔT and Δp represent the changes of temperature and pressure in the range of l, respectively. ∇T and ∇p represent the gradients of temperature and pressure at this place, respectively. Besides the local equilibrium approximation, the attenuation rate of fluctuation in each volume unit should be much faster than the rate of macroscopic change in the system, so that the motion of a large number of particles can be surely subordinated to statistics laws.

The local equilibrium approximation is a more important fundamental approximation in modern thermodynamics. It is useful in nearly all research to be transplanted from thermodynamics of equilibrium systems to thermodynamics of nonequilibrium systems. Besides temperature, pressure, and thermodynamic functions and parameters, the concept of the phase of a substance can also be taken as an example. In equilibrium thermodynamics a "phase" can be defined as a homogeneous and physically distinct part of a system which is separated from other parts of the system by definite bounding surfaces. In other words, it can be said that the complete physically and chemically homogeneous part inside a system is called a "phase", there are obvious boundaries between phase and phase, and its nature on the boundary is suddenly changed. Obviously, such a definition of phase cannot be used directly for nonequilibrium systems, in which there are usually temperature, pressure and/or concentration gradients. However, after the acceptance of the local equilibrium approximation, a gaseous system with temperature, pressure or concentration gradients can still be treated approximately as a single phase. For the reason that any local region of a gaseous system with temperature, pressure and/or concentration gradients is treated as part of a continuous single phase, such a whole gaseous system can be approximately treated as a single phase system. This approximation coincides with the ordinary concept. Otherwise, it would not be accepted by most people that a gaseous system with only a small temperature gradient were treated as a system with an infinite number of "phases". In this way, on the basis of the local equilibrium approximation, a multiphase system is still distinguishable from a single-phase system in modern thermodynamics. Therefore, the existence of nonequilibrium phase diagrams and corresponding phase lines are both possible and necessary. Discussions on the multiphase system of stationary nonequilibrium states, the corresponding nonequilibrium phase diagrams, thermodynamic coupling between irreversible processes in a multiphase nonequilibrium system, and so on, are rarely found in other books concerning nonequilibrium thermodynamics. In this book, discussions on these topics will be on going.

4.4 Calculations of Entropy Productions

According to classical thermodynamics, for isolated systems an irreversible process will result in a positive entropy change of the system, i.e. $(\mathrm{d}S)_{\mathrm{iso}} > 0$, but for closed and open systems the entropy change of the system may be positive or negative. There is no general quantitative relationship between the irreversible process and the entropy change of the system in classical thermodynamics. However, based on modern expression of the second law of thermodynamics, there are some quantitative relationships between the irreversible process and the entropy production of the system, i.e. any single practical irreversible process will result in a positive entropy production ($\mathrm{d_i}S > 0$ for any isolated system, closed system or open system). Due to the influence of classical thermodynamics, modern thermodynamics at the time of its first or exploration stage of development was called "nonequilibrium thermodynamics" or "thermodynamics for irreversible processes", and regarded as simply belonging to the inequality part of the second law of thermodynamics. Therefore, the first task for modern thermodynamics at that time was regarded as how to find equality or equalities for calculating the entropy production of irreversible processes. Such work to a certain degree had already been partially done. For the purposes of this book some resulting formulations will be listed here. The details on the sources of these formulae can be found in the references. [2], (p. 68–72) [3]

For nonequilibrium thermodynamic systems or thermodynamic systems with irreversible processes the total entropy change with time may be written as

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\int_V s\mathrm{d}V = \int_V \frac{\partial s}{\partial t}\mathrm{d}V = \int_V \mathrm{d}V\left(-\nabla\cdot J_s + \sigma\right), \tag{4.17}$$

where, s represents the entropy value per unit volume; J_s represents the entropy flow going out through unit area per unit time and is called the entropy flux; $\nabla\cdot J_s$ is the divergence of the entropy flux, which represents the entropy flowing out from unit volume per unit time and is called the velocity of the entropy flow; and σ is the entropy production in unit volume per unit time and is called the velocity of the entropy production or the strength of the entropy source. Comparison with $\mathrm{d}S = \mathrm{d_e}S + \mathrm{d_i}S$ gives

$$\frac{\partial s}{\partial t} = -\nabla\cdot J_s + \sigma. \tag{4.18}$$

The first term on the right side is the velocity of the entropy coming from the surroundings and flowing into a definite volume unit. The second term is the velocity of the entropy production due to the irreversible process inside the definite volume unit of the system.

4.4.1 For Heat Conduction

For a homogeneous isotropic system with only heat conduction, but without influence of the external field and without internal viscose convection, the increase of internal energy in volume unit is the result of heat flowing into the unit. Now u represents internal energy density of volume unit, and J_q represents heat flow flux (heat coming into the volume unit through unit area per unit time) or called thermodynamic flow (or heat conduction flow). According to the conservation law of energy one may obtain

$$\frac{\partial u}{\partial t} = -\nabla\cdot J_q. \tag{4.19}$$

For a closed volume unit with no cubage work, from thermodynamic fundamental differential Eq. (3.16) $\mathrm{d}U = T\mathrm{d}S - p\mathrm{d}V$, we obtain

$$T\mathrm{d}s = \mathrm{d}u. \tag{4.20}$$

Here, s is entropy density of the volume unit. T is independent of time, the increase velocity of entropy density is

$$\frac{\partial s}{\partial t} = \frac{1}{T}\frac{\partial u}{\partial t} = -\frac{1}{T}\nabla \cdot J_q. \tag{4.21}$$

Because

$$\frac{1}{T}\nabla \cdot J_q = \nabla \cdot \left(\frac{J_q}{T}\right) - J_q \cdot \nabla \left(\frac{1}{T}\right), \tag{4.22}$$

$$\frac{\partial s}{\partial t} = -\nabla \cdot \frac{J_q}{T} + J_q \cdot \nabla \left(\frac{1}{T}\right). \tag{4.23}$$

Here,

$$-\nabla \cdot \frac{J_q}{T} = \frac{\partial_{\mathrm{e}} s}{\partial t}. \tag{4.24}$$

That is, the increase velocity of entropy density of the volume unit is just due to heat flow coming from exterior, while $\frac{J_q}{T}$ is just the density of entropy flow.

$$J_q \cdot \nabla \left(\frac{1}{T}\right) = \frac{\partial_{\mathrm{i}} s}{\partial t} = \sigma. \tag{4.25}$$

σ is the entropy production velocity of volume unit due to the irreversible process, i.e. local entropy production velocity. If set $X_q = \nabla\left(\frac{1}{T}\right)$, and call it thermodynamic force of heat conduction (or heat conduction force), then

$$\sigma = J_q \cdot X_q. \tag{4.26}$$

Therefore, local entropy production velocity of heat conduction equals the inner production of the heat conduction flow (flux of heat flow density) and the heat conduction force.

The inhomogeneous temperature causes heat conduction. Fourier law of heat conduction is that flux of heat flow density (heat conduction flow) is positively proportional to the gradient of temperature, i.e.

$$J_q = -\kappa \nabla T. \tag{4.27}$$

Here, κ is the coefficient of heat conduction. Substitute it into Eq. (4.25),

$$\sigma = -\kappa \nabla T \cdot \nabla \left(\frac{1}{T}\right) = \kappa \frac{(\nabla T)^2}{T^2} > 0. \tag{4.28}$$

The coefficient of heat conduction κ is always bigger than zero, so $\sigma > 0$. It means that heat conduction from high temperature to low temperature is a spontaneous irreversible process, and its local entropy production velocity is always bigger than zero.

4.4.2 For Heat Conduction together with Matter Transport

For a volume unit of a single component system, its concentration change accords with continuous equation.

$$\frac{\partial n}{\partial t} + \nabla \cdot J_n = 0. \tag{4.29}$$

Here, J_n is the mass flow density, i.e. mole number of matter coming into volume unit through unit area per unit time. Equation (4.29) is the mathematical expression of conservation law of matter. Meanwhile, the mathematical expression of conservation law of energy is

$$\frac{\partial u}{\partial t} + \nabla \cdot J_u = 0. \tag{4.30}$$

Here, u is internal energy density of matter in a volume unit, and J_u is vector of energy flow density, i.e. energy coming into volume unit through unit area per unit time. Equation (4.30) is the continuous equation for energy density of the volume unit, i.e. energy increase of volume unit per unit time equals energy flow coming from exterior.

From thermodynamic fundamental differential Eq. (3.84) $\mathrm{d}U = T\mathrm{d}S - p\mathrm{d}V + \sum_i \mu_i \mathrm{d}n_i$ it is known that for a volume unit of a single component system without work exchange with exterior, the discussed Eq. (3.84) of volume unit may be rewritten into $T\mathrm{d}s = \mathrm{d}u - \mu\mathrm{d}n$. Here, the increase of matter density of volume unit $\mathrm{d}n$ will result in the increase of the internal energy density $\mu\mathrm{d}n$. Here, s, u and n are entropy density, internal energy density and matter density, respectively. μ is chemical potential of 1 mole matter. Therefore, during the simultaneous existences of heat flow and mass flow the vector of energy flow density equals

$$J_u = J_q + \mu J_n. \tag{4.31}$$

That is, the vector of energy flow density equals the sum of the vectors of energy flow density brought by the heat flow density and the mass flow density. Substitute this equation for the previous Eq. (4.30) then we get

$$\frac{\partial u}{\partial t} = -\nabla \cdot J_q - \nabla \cdot (\mu J_n). \tag{4.32}$$

Therefore, the entropy density increase velocity of the volume unit is

$$\frac{\partial s}{\partial t} = \frac{1}{T}\frac{\partial u}{\partial t} - \frac{\mu}{T}\frac{\partial n}{\partial t}. \tag{4.33}$$

Substituting Eq. (4.29) $\frac{\partial n}{\partial t} + \nabla \cdot J_n = 0$ and Eq. (4.32) for this equation, and due to $\frac{1}{T}\nabla \cdot J_q = \nabla \cdot \left(\frac{J_q}{T}\right) - J_q \cdot \nabla \left(\frac{1}{T}\right)$ and $\frac{\mu}{T}\nabla \cdot J_n - \frac{1}{T}\nabla \cdot (\mu J_n) = -\frac{1}{T} J_n \cdot \nabla\mu$, we obtain

$$\begin{aligned} \frac{\partial s}{\partial t} &= -\frac{1}{T}\nabla \cdot J_q - \frac{1}{T}\nabla \cdot (\mu J_n) + \frac{\mu}{T}\nabla \cdot J_n \\ &= -\nabla \cdot \left(\frac{J_q}{T}\right) + J_q \cdot \nabla \left(\frac{1}{T}\right) - \frac{1}{T} J_n \cdot \nabla\mu. \end{aligned} \tag{4.34}$$

The first term on the right side in Eq. (4.34) is the local entropy increase velocity due to heat flow coming into the volume unit. The second term is the local entropy increase velocity due to heat conduction of the temperature gradient in the volume unit. The third term is the local entropy increase velocity due to diffusion of the matter concentration gradient in the volume unit. If set $X_n \equiv -\frac{1}{T}\nabla\mu$, and call it thermodynamic force of mass flow, then the third term becomes $J_n \cdot X_n$. According to Eq. (4.25), $J_q \cdot \nabla\left(\frac{1}{T}\right) = \frac{\partial_e s}{\partial t} = \sigma$, and we obtain

$$\sigma = \frac{\partial_i s}{\partial t} = J_q \cdot X_q + J_n \cdot X_n. \tag{4.35}$$

and

$$\frac{\partial_e s}{\partial t} = -\nabla \cdot \left(\frac{J_q}{T}\right). \tag{4.36}$$

4.4.3 General Expressions for Entropy Production Calculations

From the expression of σ, Eq. (4.35) it can readily be shown in the formula for σ that each term on the right side is always a product of two factors. The factor that is related to the velocity of irreversible processes can in general be called the thermodynamic flow, such as heat flow, mass flow etc. and denoted by the symbol J_k. Another factor, which is related to the driving force of the irreversible thermodynamic flow, can in general be called the thermodynamic force driving the irreversible process, such as temperature gradient, concentration gradient etc., and can be denoted by X_k. Thermodynamic force is the cause of thermodynamic flow; while thermodynamic flow is the result of thermodynamic force. Now the formula for σ can simply be written in its general form:

$$\sigma = \sum_k J_k X_k. \tag{4.37}$$

Equation (4.37) is of general meaning, i.e. the local entropy production velocity of multi-irreversible-process systems equals the sum of every inner production of the vector of thermodynamic flow and the vector of thermodynamic force.

According to the early study of nonequilibrium thermodynamics, based on Eq. (4.18) $\frac{\partial s}{\partial t} = -\nabla \cdot J_s + \sigma$, the detailed expressions of entropy flux J_s and entropy production velocity σ for more complex multi-process systems can be written as

$$J_s = su + \frac{j_q}{T} - \sum_i \frac{\mu_i}{T} j_i \tag{4.38}$$

and

$$\sigma = j_q \cdot \nabla\left(\frac{1}{T}\right) + \sum_i j_i \cdot \left[-\nabla\frac{\mu_i}{T} + \frac{M_i F_i}{T}\right] - \frac{1}{T}\Pi : \nabla u + \sum_\rho \frac{A_\rho}{T} r_\rho. \tag{4.39}$$

Here,

$$A_\rho = -\sum_i \nu_{i\rho}\mu_i. \tag{4.40}$$

M_i is the molecular weight of component i, F_i is the force on unit amount of component i due to the external field, Π is the tensor of the viscosity stress. A_ρ is the affinity of the ρth reaction, and r_ρ is the reaction rate of the ρth reaction. μ_i is chemical potential of reaction component i, while $\nu_{i\rho}$ is the stoichiometric coefficient in chemical reaction ρ of component i.

Each term in both Eq. (4.38) for J_s and Eq. (4.39) for σ has clear physical meanings. The first term on right side in the formula for J_s is the entropy flux due to convection, the second term is the entropy flux due to heat conduction, and the third term is the entropy flux due to diffusion. Here, u is the velocity of the mass center of a fluid volume unit; j_q is the heat through unit area per unit time, i.e. the flux of the heat flow; and j_i is the flux of the mass flow of component i through unit area per unit time. The first term on the right in the formula for σ is concerned with heat conduction, the second term is concerned with diffusion process (including natural diffusion and diffusion under external field), the third term is concerned with viscous flow, and the fourth term is concerned with the chemical reaction. For reference, some detailed expressions of thermodynamic flows and thermodynamic forces of the irreversible processes are listed in Table 4.1.

Table 4.1 Flows J_k and forces X_k of irreversible processes

Irreversible process	J_k	X_k
Diffusion	Mass flow j_i	$-\nabla\frac{\mu_i}{T}+\frac{M_iF_i}{T}$
Heat conduction	Heat flow j_q	$\nabla\left(\frac{1}{T}\right)$
Viscose flow	Tensor of viscosity stress Π/T	$-\nabla u$
Chemical reaction	Reaction rate r_ρ	$\frac{A_\rho}{T}$

It should be pointed out that the forces and the flows of the irreversible processes defined and listed in Table 4.1 are not a unique set of the combination between J_k and X_k. There are some arbitrary sets of combination to be selected. However, once a set of formulations of forces or flows (e.g. forces at first) has been decided, then another set of formulations of flows or forces (e.g. flows) in form are also decided in company. Any decisions should accord with the following requirements:

1. The product of the force and the flow in any set should have the dimension of entropy production.
2. For the same system and the same set of irreversible processes, the sum of the products should remain constant with respect to the formulation changes of forces and flows. That is to say, if a new set of formulations of forces X'_k are defined, the new set should be a linear combination of the original formulations of forces X_k, and the entropy productions should remain constant for the selection of a new set of formulations of flows J'_k. That is,
$$\sum_k J_k X_k = \sum_k J'_k X'_k. \tag{4.41}$$

It may be found from different references that the driving force (or called the affinity) of chemical reactions is sometimes written as A and sometimes A/T for the reasons discussed. It should be noted that the definition of affinity is sometimes written as $A = -\sum_\rho \nu_{i\rho}\mu_i$, and sometimes the definition is written as $A = -\frac{1}{T}\sum_\rho \nu_{i\rho}\mu_i$. It is important that the dimension of the product of a force and a flow must be the same as the dimension of entropy production.

The introduction of the entropy production $\mathrm{d_i}S$ or the velocity of the entropy production σ is very important for developing modern thermodynamics. In some respects its role is similar to the role of the entropy function in classical thermodynamics. For a clearer understanding, the dissipative function Φ can also be adopted. Here, $\Phi = T\sigma$, and Φ has the dimension of a free energy in unit volume per unit time. It can be used to measure the dissipation rate of the free energy due to irreversible processes, so it can be called the dissipation function.

4.5 Thermodynamic Coupling of Modern Thermodynamics

Thermodynamic coupling is the watershed between modern thermodynamics and classical thermodynamics. Thermodynamic coupling is also the core of modern thermodynamic and the most important fundamental concept of modern thermodynamics. Thermodynamic coupling has come to be clearly understood after suffering different kinds of difficulties since the birth of thermodynamics. Therefore, the fundamental discussions on thermodynamic coupling will firstly be emphasized in this section.

The phenomenon of thermodynamic coupling could be naturally concluded on the basis of the second law of thermodynamics after its mathematical expression was obtained. As early as 1865, in Clausius' famous paper introducing entropy function it was said:(p. 364)[11]

> The second fundamental theorem, in the form which I have given to it, asserts that all transformations occurring in nature may take place in a certain direction, which I have assumed as positive, by themselves, that is, without compensation; but that in the opposite, and consequently negative direction, they can only take place in such a manner as to be compensated by simultaneous occurring positive transformations.

In fact, the words, "all transformations occurring in nature may take place in a certain direction, which I have assumed as positive, by themselves, that is, without compensation", here described simple systems with only spontaneous processes without thermodynamic coupling (which we now call uncoupling systems in classical thermodynamics); while the words, "in the opposite, and consequently negative direction, they can only take place in such a manner as to be compensated by simultaneous occurring positive transformations", described complex systems including simultaneous nonspontaneous process and spontaneous process with thermodynamic coupling (which we now call coupling systems in modern thermodynamics). The Clausius' "compensation" is now called thermodynamic coupling. However, Clausius' words cited here did not attract sufficient attention for a long time.

During the time approaching the 20th century, thermodynamic researches on simple systems had been rather systematic and complete, and complex systems were also under consideration. Just as mentioned in the last section, many physicists paid attention to quantitative calculations of entropy production, while in contrast chemists and biochemists focused attention to macroscopic phenomena and the summarization of experiences. The first and the second laws of thermodynamics were just based on the summarization of experiences. In 1931, biochemist Dean Burk, directly on the basis of the mathematical expression of the second law of thermodynamics, clearly introduced an important conception of "coupled reactions", or what is called "thermodynamic coupling" in his paper, "*The Reversibility of Coupled Reactions in Biological Systems and The Second Law of Thermodynamics*".[12] He also wrote out the mathematical expression of thermodynamic coupling in Gibbs free energy form. If Gibbs free energy changes of two (two kinds of) reactions and of the system (or the whole reaction) accorded with the following equation set, then the isothermal isobaric reaction system was called coupled reactions. In the 1930s, the function of Gibbs free energy was usually represented by F, [13]

$$\Delta F_1 > 0,\ \Delta F_2 < 0\ \&\ \Delta F \leqslant 0. \tag{4.42}$$

In his paper, Burk used the equation $\Delta F = (\Delta F_1 + \Delta F_2)$. ΔF in chemistry fora usually represents the Gibbs free energy change after finishing a chemical reaction equation formula, so the time intervals for reaction 1 and reaction 2 to get ΔF_1 and ΔF_1, respectively, may not be the same. Therefore, $\Delta F = \Delta F_1 + \chi \Delta F_2$ should be written. Here, $\chi = r_2/r_1$ is the rate ratio of two reactions, but in Burk's paper the reaction rate ratio was regarded as 1. The factor of rate ratio was out of his consideration. Now the criterion set for thermodynamic coupling for isothermal isobaric systems had better be written as $[(\Delta G_1)_{T,p} > 0,\ (\Delta G_2)_{T,p} < 0\ \&\ (\Delta G)_{T,p} \leqslant 0]$. This is just the same as Eq. (4.13). Here, $(\Delta G_1)_{T,p}$, $(\Delta G_2)_{T,p}$ and $(\Delta G)_{T,p}$ are molar Gibbs free energy changes of nonspontaneous reaction (denoted by subscript 1), spontaneous reaction (denoted by subscript 2) and the system (the whole reaction), respectively, at constant temperature and pressure and $(\Delta G)_{T,p} = (\Delta G_1)_{T,p} + \chi (\Delta G_2)_{T,p}$ ($\chi = r_2/r_1$ is the reaction rate ratio of reaction 2 to reaction 1).

In comparison with the Eq. (3.103) in Chapter 3, i.e. $A = -\Delta G_\mathrm{m}$, which was defined by Th. de Donder of the Brussels school, the criterion of thermodynamic coupling, Eq. (4.7) $[\mathrm{d_i}S_1 < 0,\ \mathrm{d_i}S_2 > 0\ \&\ \mathrm{d_i}S \geqslant 0]$, could also written in its chemical affinity form, [14]

$$A_1 r_1 < 0,\ A_2 r_2 > 0,\ \&\ A_1 r_1 + A_2 r_2 \geqslant 0, \tag{4.43}$$

but the equality of this expression set in Prigogine's book was sometimes left forgotten. For instance, Prigogine wrote: (p. 25) [2]

> The second principle requires that the entropy production resulting from all the simultaneous reactions is positive. However, it may happen that a system undergoes two simultaneous reactions such that
>
> $$A_1 r_1 < 0,\ A_2 r_2 > 0 \tag{4.44}$$
>
> provided that the sum
>
> $$A_1 r_1 + A_2 r_2 > 0. \tag{4.45}$$

Although Prigogine correctly wrote that (p. 25) [2]

> Both reactions are then called 'coupled' reactions. Thermodynamic coupling allows one of the reactions to progress in a direction contrary to that prescribed by its own affinity.

Because at that time it was very popular that modern thermodynamics was called "nonequilibrium thermodynamics", and was regarded as belonging to the inequality of the second law of thermodynamics. Later we can see that something similar to such a missing of the equality part of the second law of thermodynamics can also be found in Onsager's papers.

However, as early as in 1937, an American visiting scientist of De Donder's group, P. van Rysselberghe, coming from Stanford University wrote in his published paper that [15, 16], (p. 50–51) [17]

> It seems fairly certain that coupling of a reaction with negative affinity by a reaction with a positive affinity is possible only if at least one reactant or product is common to both reactions.

and called such “common” substance as “coupling substance”. In this way, a suspect premise, with so-called “seems fairly certain”, was added into the criterion of thermodynamic coupling. That is, a “coupling substance” would be necessary to be found for thermodynamic coupling. Later, from Prigogine’s paper, in which van Rysselberghe’s “example of thermodynamic coupling” was cited, it can be found that another condition of “under standard conditions” had also been included. So the nature of the fundamental concept of thermodynamic coupling was changed. At the same time, a lot of “examples of thermodynamic coupling” which were originally difficult to find in ordinary chemistry suddenly emerged. Because thermodynamics data “under standard conditions” can easily be found from thermodynamics databooks, this simple way with no necessity for experimental work seemed to be “welcome” also in biochemistry. During 1870s–1980s, such a revised “thermodynamic coupling” concept became a fashion in the ordinary (inanimate) chemistry. That had the important fundamental concept of thermodynamic coupling arrive at a crossroads because the criterion of the revised “thermodynamics coupling” changed Eq. (4.43) into

$$A_1^0 r_1 < 0,\ A_2^0 r_2 > 0,\ \&\ A_1^0 r_1 + A_2^0 r_2 \geqslant 0. \tag{4.46}$$

Here, all “chemical affinities” with superscript (^{0}) were the value under standard conditions. For the most useful criterion of thermodynamic coupling for isothermal isobaric systems Eq. (4.13) would also be changed into

$$\left(\Delta G_1^0\right)_{T,p} > 0,\ \left(\Delta G_2^0\right)_{T,p} < 0\ \&\ \left(\Delta G^0\right)_{T,p} \leqslant 0. \tag{4.47}$$

In this way, in the rigorous field of thermodynamics two quite different kinds of concepts ΔG^0 and ΔG were actually confused with each other, and then “thermodynamic coupling” became a fashionable label in chemistry for examples of the mass action law, Le Chatelier principle or equilibrium point shift.

For instance, Prigogine had cited van Rysselberghe’s so-called “demonstration” of reaction coupling “under standard conditions” in his book, “*Chemical Thermodynamics*”, and considered: (p. 42) [18]

> Thermodynamics coupling is an essential feature of living systems. For example the synthesis of urea is coupled with the combustion of glucose and other substances in the liver. This synthesis may be represented schematically by [3]
>
> $$2\,\mathrm{NH_3} + \mathrm{CO_2} \rightarrow (\mathrm{NH_2})_2\mathrm{CO} + \mathrm{H_2O}\ \text{(coupled reaction)}, \tag{4.48}$$
>
> $$\frac{1}{8}\mathrm{C_6H_{12}O_6} + \mathrm{O_2} \rightarrow \mathrm{CO_2} + \mathrm{H_2O}\ \text{(coupling reaction)} \tag{4.49}$$
>
> The affinities of these reactions under standard conditions are found to be
>
> $$A_1 = -11{,}000\ \text{cal}. \tag{4.50}$$
>
> $$A_2 = +115{,}000\ \text{cal}. \tag{4.51}$$

In fact, the phrase “under standard conditions” here means that A_1 and A_2 here should be represented by A_1^0 and A_2^0, respectively. Obviously, this ambiguous change of the criterion of reaction coupling into its incorrect form, i.e. Eq. (4.46), had already been accepted by

---

[3] Here and hereafter numbers of cited equations are added or revised as the style of this book

Prigogine at that time.

Up to now, this seems to have been adopted in all textbooks of bioenergetics and biochemistry which we have found. That is, the incorrect criterion of reaction coupling, Eq. (4.47), was also practically accepted or written as follows.

$$\left(\Delta G_1'^0\right)_{T,p} > 0,\ \left(\Delta G_2'^0\right)_{T,p} < 0 \ \&\ \left(\Delta G'^0\right)_{T,p} \leqslant 0, \tag{4.52}$$

where, $\left(\Delta G'^0\right)_{T,p}$ = $\left(\Delta G_1'^0\right)_{T,p}$ + $\left(\Delta G_2'^0\right)_{T,p}$ and $\left(\Delta G'^0\right)_{T,p}$ represents the change of the Gibbs free energy of a reaction taking place in a solution in which all the reactants and products are in the standard state [The standard state of a substance in solution is defined as a solution of substance at a concentration of 1 M, i.e. 1 mol kg^{-1} or 1 mol dm^{-3} (if the H$^+$ ion participates in a reaction, then the pH value equals 7) and at 25 °C]. χ =1 is also unsuitably adopted. In this way ΔG incorrectly equal to ΔG^0 or $\Delta G'^0$ with the same sign has been assumed. It can be generally found that in textbooks of physical chemistry, biochemistry and other related books, the description of these kinds of biochemical reactions are often similar:

$$\mathrm{ATP} + \mathrm{H_2O} \rightarrow \mathrm{ADP} + \mathrm{P_i} \qquad \Delta G'^0 = -30.5\mathrm{kJ}. \tag{4.53}$$

$$\mathrm{glucose} + \mathrm{Pi} \rightarrow \mathrm{glucose} - 6 - \mathrm{phosphate} + \mathrm{H_2O} \qquad \Delta G'^0 = +13.8\mathrm{kJ}. \tag{4.54}$$

$$\mathrm{ATP} + \mathrm{glucose} \rightarrow \mathrm{ADP} + \mathrm{glucose} - 6 - \mathrm{phosphate} \qquad \Delta G'^0 = -16.7\mathrm{kJ}. \tag{4.55}$$

In some books, such as in Southerland's "*Biochemistry*", there was a notation that [19]

> Usually, ΔG for biologic reactions has the same sign as ΔG^0, and the two often do not differ greatly in value,

but such an "approximation" was still incorrect. After Prigogine won the 1977 Nobel Prize, the terms "nonequilibrium thermodynamic coupling", "reaction coupling" and "coupling reaction" became fashionable. The "approximate" or incorrect criteria, Eq. (4.46), Eq. (4.47) or Eq. (4.52), were also introduced into the inanimate chemistry field. However, in the inanimate chemistry field such kinds of mistakes were more obvious. All of examples of the so-called "reaction coupling" or "coupled reactions" proposed during 1970s – 1980s in ordinary chemistry were wrong. For instance, an example of "coupled reactions" often cited by other textbooks of physical chemistry or in the literature is in the paragraph "Coupled Reactions" in the book, "*Physical Chemistry*" written by Castellan. The passage reads: [5, 20]

> 11.18 *COUPLED REACTIONS*$\cdots$
>
> The reaction scheme is coupled reactions
>
> $$\mathrm{TiO_2(s)} + 2\mathrm{Cl_2(g)} \rightarrow \mathrm{TiCl_4(l)} + \mathrm{CO_2(g)}, \quad \Delta G_1^0(298\mathrm{K}) = 152.3\ \mathrm{kJ\ mol^{-1}}, \tag{4.56}$$
>
> $$\mathrm{C(s)} + \mathrm{O_2(g)} \rightarrow \mathrm{CO_2(g)}, \quad \Delta G_2^0(298\mathrm{K}) = -394.4\ \mathrm{kJ\ mol^{-1}}, \tag{4.57}$$
>
> and the overall reaction is

> $$C(s) + TiO_2(s) + 2Cl_2(g) \rightarrow TiCl_4(l) + CO_2(g), \qquad (4.58)$$
>
> $$\Delta G^0(298\text{K}) = -242.1 \text{ kJ mol}^{-1}.$$
>
> $$\Delta G^0(298\text{K}) = \Delta G_1^0(298\text{K}) + \Delta G_2^0(298\text{K}) < 0. \qquad (4.59)$$
>
> Since the overall reaction has a highly negative ΔG^0, it is spontaneous. As a general rule metal oxides cannot be converted to chlorides by simple replacement; in the presence of carbon, the chlorination proceeds easily. Coupled reactions have great importance in biological systems$\cdots$.

Note that ΔG^0 had been unsuitably used as a criterion of a spontaneous reaction. The words "Since the overall reaction has a highly negative ΔG^0, it is spontaneous." are obviously incorrect. In fact, the overall reaction practically occurs at about 1200°C and 100 kPa. Under conditions of the existence of white-hot carbon the concentration or partial pressure of O_2 is very low. The partial pressure of O_2 is absolutely impossible to be a standard pressure. Therefore, the data of ΔG_1^0 and ΔG_2^0 are meaningless as a criterion for the practical conditions, while ΔG_1 and ΔG_2 of these two so-called "coupled reactions" are practically of small negative value, i.e. $\Delta G_1 < 0$ and $\Delta G_2 < 0$. Both of these two reactions are spontaneous toward the right side, and there is nothing concerned with thermodynamic coupling at all. The introduction of carbon at high temperature is only for removing the product O_2 of the reaction between TiO_2 and Cl_2, and the reaction is going toward the right side. That is the well-known La Chatelier principle, the mass action law, the equilibrium shift principle, or also called simultaneous equilibrium.

Now in some thermodynamics books, such as in "*Modern Thermodynamics*" written by D. Kondepudi and I. Prigogine,[1] the chain reaction of the synthesis of HBr from H_2 and Br_2 was also unsuitably called "reaction coupling" or "coupled reactions".

From thermodynamic data books it is easy to find the standard Gibbs free energy values of all concerned species of the synthesis of HBr from H_2 and Br_2. The reaction usually occurs at the temperature between 500 K and 600 K, so the values of standard Gibbs free energy, i.e. ΔG^0 values at 500 K and at 600 K together with the values at 298 K, are listed.

At the beginning of synthesis of hydrogen bromide

$$Br_2 = 2Br, \qquad (4.60)$$

$$\Delta G^0_{298\text{K}} = 161.72 \text{ kJ mol}^{-1},$$

$$\Delta G^0_{500\text{K}} = 140.35 \text{ kJ mol}^{-1},$$

$$\Delta G^0_{600\text{K}} = 129.63 \text{ kJ mol}^{-1}.$$

After the formation of chain reaction, in the first step,

$$Br_2 + H_2 = Br + H, \qquad (4.61)$$

$$\Delta G^0_{298\text{K}} = 61.37 \text{ kJ mol}^{-1},$$

$$\Delta G^0_{500\text{K}} = 65.81 \text{ kJ mol}^{-1},$$

$$\Delta G^0_{600\text{K}} = 65.01 \text{ kJ mol}^{-1}.$$

In the second step,

$$Br_2 + H = HBr + Br, \qquad (4.62)$$

$$\Delta G^0_{298\text{K}} = -177.49 \text{ kJ mol}^{-1},$$

$$\Delta G^0_{500\text{K}} = -179.80 \text{ kJ mol}^{-1},$$

$$\Delta G^0_{600\text{K}} = -180.69 \text{ kJ mol}^{-1}.$$

The overview reaction is

$$\begin{aligned} & \mathrm{Br_2 + H = HBr + Br}, \qquad (4.63) \\ & \Delta G^0_{298\mathrm{K}} = -110.12\ \mathrm{kJ\ mol^{-1}}, \\ & \Delta G^0_{500\mathrm{K}} = -113.99\ \mathrm{kJ\ mol^{-1}}, \\ & \Delta G^0_{600\mathrm{K}} = -115.68\ \mathrm{kJ\ mol^{-1}}. \end{aligned}$$

It meant that the incorrect criterion of thermodynamic coupling, i.e. Eq. (4.47) [$\left(\Delta G^0_1\right)_{T,p} > 0$, $\left(\Delta G^0_2\right)_{T,p} < 0$ & $\left(\Delta G^0\right)_{T,p} \leqslant 0$] was used by Kondepudi and Prigogine, and the synthesis of HBr from H_2 and Br_2 was unsuitably called “reaction coupling” or “coupled reactions”.

Besides, many other reactions, such as the dehydrogenation of organic compounds being promoted by adding small amount of oxygen, were also called “reaction coupling” at that time. In a textbook of “*Physical Chemistry*” (p. 427–431, p. 437–444)[5], it was written that

> For instance, the dehydrogenation of ethylbenzene is
>
> $$\text{(a) } \mathrm{C_8H_{10}\ (g) = C_8H_8 + H_2\ (g)}, \quad \text{at 298 K}, K^0_{p\ 298\mathrm{K}} = 2.7 \times 10^{-15}. \qquad (4.64)$$
>
> The oxidation of ethylbenzene results in dehydration.
>
> $$\text{(b) } \mathrm{C_8H_{10}\ (g) + 0.5\ O_2(g) = C_8H_8 + H_2O\ (g)}, \quad \text{at 298 K}, K^0_{p\ 298\mathrm{K}} = 2.9 \times 10^{25}. \qquad (4.65)$$
>
> Therefore, in reaction (a) nearly no observable styrene could appear, but reaction (b) can completely produce styrene. Try to make an analysis on these reactions.
>
> $$\text{(c) } \mathrm{H_2\ (g) + 0.5\ O_2(g) = H_2O\ (g)}. \qquad (4.66)$$
>
> For reaction (c) $\Delta G^0_{298\ \mathrm{K}} = -228.59\ \mathrm{kJ\ mol^{-1}}$, $K^0_{p\ 298\mathrm{K}} = 1.26 \times 10^{40}$, so reaction (b) may be regarded as coupled reactions consisting of reaction (a) and reaction (c).

Moreover,

> ···The production of acrylonitrile from propene reaction may be another example.
>
> $$\mathrm{CH_2 = CHCH_2 + NH_3 = CH_2 = CHCN + 3\ H_2}. \qquad (4.67)$$
>
> The yield of acrylonitrile from ammoniation is low but the yield of acrylonitrile from ammonio-oxidation is rather high.
>
> $$\mathrm{CH_2 = CHCH_2 + NH_3 + 1.5\ O_2 = CH_2 = CHCN + 3\ H_2O}. \qquad (4.68)$$
>
> That may be regarded as coupled reactions consisting of the former reaction and $3\ H_2 + 1.5\ O_2 = 3\ H_2O$.

All such kinds of problems were a consequence of using the incorrect criterion of Eq. (4.47) etc. and should be corrected.

The confusion between “equilibrium shift” and “reaction coupling” had already been felt to a certain extent even at that time. Therefore, two parallel definitions of “reaction coupling" could be found in some textbooks, although that was still incorrect. For instance, Leidler and Meiser in their book "*Physical Chemistry*" wrote [21]

> 4.7 Coupling of Reactions
>
> For a chemical reaction
>
> 1.
>
> $$A+B \longleftrightarrow X+Y \tag{4.69}$$
>
> removal of either of the products X or Y will lead to a shift of equilibrium to the right. Frequently a product is removed by the occurrence of another reaction; for example, X might isomerize into Z.
>
> 2.
>
> $$X \longleftrightarrow Z \tag{4.70}$$
>
> We can write an equilibrium constant for reaction 1,
>
> $$K_1 = \left(\frac{[X]\,[Y]}{[A]\,[B]}\right)_{\text{eq}} \tag{4.71}$$
>
> and a corresponding standard Gibbs energy change, $\Delta G_1^0 = -RT\ln K_1$. Similarly for the second reaction
>
> $$K_2 = \left(\frac{[Z]}{[X]}\right)_{\text{eq}} \tag{4.72}$$
>
> and $\Delta G_2^0 = -RT\ln K_2$. We can add reaction 1 and 2 to obtain
>
> 3.
>
> $$A+B \longleftrightarrow X+Z \tag{4.73}$$
>
> The equilibrium constant K_3 for the combined reaction 3 is
>
> $$K_3 = \left(\frac{[Y]\,[Z]}{[A]\,[B]}\right)_{\text{eq}}. \tag{4.74}$$
>
> The same result is obtained by multiplying K_1 and K_2.
>
> $$K_1K_2 = \left(\frac{[X]\,[Y]}{[A]\,[B]}\right)_{\text{eq}} \times \left(\frac{[Z]}{[X]}\right)_{\text{eq}} = \left(\frac{[Y]\,[Z]}{[A]\,[B]}\right)_{\text{eq}} = K_3. \tag{4.75}$$
>
> We can take natural logarithms of both sides of this equation, and then multiply by $-RT$,
>
> $$-RT\ln K_3 = -RT\ln K_1 - RT\ln K_2 \tag{4.76}$$
>
> so that
>
> $$\Delta G_3^0 = \Delta G_1^0 + \Delta G_2^0. \tag{4.77}$$
>
> The result is easily shown to be quite general; if we add reactions, the ΔG^0 of the resulting reaction is the sum of the ΔG^0 values of the component reactions$\cdots$.
>
> As a result of this coupling of chemical reactions, it is quite possible for a reaction to occur to a considerable extent even though it has a positive value of ΔG^0. Thus for the scheme just given, suppose that reaction 1 has a positive value, $\Delta G_1^0 > 0$. Reaction 1 by itself will therefore not occur to a considerable extent ($K_1 > 1$). Suppose, however, that reaction 2 has a negative value. It is then possible for ΔG_3^0 to be negative, so that A + B may react to a considerable extent to give X+Y. In terms of the Le Chatelier principle, this simply means that X is constantly removed by the isomerization reaction 2, so that the equilibrium of reaction 1 is shifted to the right.

Leidler and Meiser continue:

> There is another way in which chemical reactions can be coupled. In the cases we have considered, the product of one reaction is removed in a second reaction. Alternatively, two reactions might be coupled by a catalyst. Suppose, for example, that we have two independent reactions, having no

> common reactants or products:
> 1.
> $$A + B \longleftrightarrow X + Y \tag{4.78}$$
> 2.
> $$C \longleftrightarrow Z \tag{4.79}$$
> There might exist some catalyst that brings about the reaction 3.
> $$A + B + C \longleftrightarrow X + Y + Z \tag{4.80}$$
> and this catalyst would couple the reactions 1 and 2. This type of coupling of reactions is quite common in biological systems.

In that book the incorrect descriptions were based two kinds of so-called "definitions" of "reaction coupling". Similar incorrectness of "reaction coupling" is often found in books of chemical thermodynamics.

In 1983, M. Boudart in his paper "*Thermodynamic and Kinetic Coupling of Chain and Catalytic Reactions*" proposed his objections. [22, 23] He wrote:

> Any guidance from thermodynamics in the art and science of chemical kinetics has a value which is particularly appreciated as it is limited but general. An intriguing example of such thermodynamic assistance to chemical kinetics comes from the De Donder's inequality $Ar \geqslant 0$, linking the affinity $A = -\Delta G$ of a reaction and its rate r. Indeed, if two reactions (subscripts 1 and 2) occur simultaneously, it is conceivable that the following inequalities $A_1 r_1 < 0$, $A_2 r_2 > 0$ might hold provided that $A_1 r_1 + A_2 r_2 > 0$. It is said that the first reaction is coupled to the second one, 'thermodynamics coupling thus allows the coupled reactions to proceed in a direction opposite that dictated by its affinity'. (p. 38) [18] Unfortunately, the coupling mechanism, clearly a notion from chemical kinetics, needs to be known before the inequality $A_1 r_1 < 0$ can be accepted. It is quite possible that, when the coupling mechanism is understood, the need for the thermodynamic coupling as defined above disappears, together with the uncomfortable inequality $A_1 r_1 < 0$, which is purely formal in the sense that the coupled reaction does not take place at all but is only a stoichiometric statement devoid of kinetic (i.e. r_1) meaning.

Boudart continued discussing examples of the so-called "thermodynamic reaction coupling" found in the literature. The first example was a chain reaction. Note that the symbol for the rate v has been changed into r to fit this book style.

> Consider first the two gas-phase reactions
> $$H_2 = 2\,H \tag{4.81}$$
> $$2\,H_2 + O_2 = 2\,H_2O \tag{4.82}$$
> taking place at 1000 K between reactants and products at their standard state (1 atm = 101.3 kPa). The standard affinity for the first one is $A_1^0 = -331$ kJ mol^{-1}, while $A_2^0 = 385$ kJ mol^{-1}. For the overall reaction
> $$3\,H_2 + O_2 = 2\,H_2O + 2\,H \tag{4.83}$$
> The standard affinity is $385 - 331 = 54$ kJ mol^{-1}, so that it would appear that reaction 1 is the reaction coupled by reaction 2 the coupling reaction, and that indeed, for reaction 1, $A_1 r_1 < 0$. But this is not so because, in fact, reaction 1 does not take place at all. Rather, according to Semenov, the branching chain reaction propagates through three elementary processes as follows:
> $$OH + H_2 \rightarrow H_2O + H \tag{4.84}$$

$$H + O_2 \rightarrow OH + O \tag{4.85}$$

$$O + H_2 \rightarrow OH + H \tag{4.86}$$

> If we sum up these three elementary processes, taking the first one twice, so that its stoichiometric number σ is given a value equal to 2, we reproduce the overall reaction 3. The overall De Donder's inequality applies to it, but 'reaction 1' has nothing to do with what happens at the mechanistic level. Indeed, reaction 1 does not take place at all. There is no thermodynamic coupling in the sense mentioned above. The multiplication of hydrogen atoms is due not to reaction 1 but to the branching chain reaction $H + O_2 = OH + O$. Thus, the concept of thermodynamic coupling disappears at the kinetic level of interpretation.

Boudart had also pointed out that the synthesis of HBr from H_2 and Br_2 is only a chain reaction in kinetics and there was nothing concerned with thermodynamic coupling. It is obvious that Boudart's disagreements are reasonable. Boudart's paper was mainly on kinetics and was not concerned with the nature of thermodynamic coupling. Note that the so-called "kinetic coupling" introduced by Boudart was also incorrect. At the same time, Boudart did not find what was the root of these mistakes of so-called "thermodynamic coupling".

In 1989, Chinese scientist R.-S. Li, who had been an outstanding doctoral graduate student of Prigogine, also lodged an objection against the abuse of the conception of reaction coupling or thermodynamic coupling.[24] After his derivation he pointed out that

> the thermodynamic coupling is impossible for those reactions which obey kinetically the mass action law.

In this way, not only in Leidler and Meiser's and Castellan's books but also in all other relevant ordinary chemistry textbooks at that time, examples of "reaction coupling" were incorrect as well.[19, 21, 25] Neither did he find what was the root of these mistakes of so-called "thermodynamic coupling".

It would be unsuitable to simply ascribe these mistakes to a misunderstanding of the difference between ΔG and ΔG^0. The biggest bepuzzlement was that no example which agreed with the criterion of thermodynamic coupling had practically been found in inanimate systems at that time. The example of the activated low-pressure diamond growth from the vapor phase now fills this gap.[26~46] After making clear the reaction mechanism of the activated low-pressure diamond growth, quantitatively calculations can be made for the activated low-pressure diamond growth. The quantitatively calculated results show that the low-pressure transformation from graphite to diamond with the association of superequilibrium atomic hydrogen accords with the mathematical expression of the second law of thermodynamics. Such quantitative demonstration not only removes more than 30 years of suspicion about low-pressure diamond growth "violating" thermodynamics, but also removes more than 60 years of mistakes on understanding reaction coupling and negative entropy production reactions.

Moreover, in this Section, "4.5 Thermodynamic Coupling of Modern Thermodynamics", it should also be pointed out that just in the same year (1931) when Burk proposed the concept of "coupled reactions" (i.e. thermodynamic coupling), Onsager also proposed a similar concept of "coupled irreversible processes", an approximate interaction between different processes in multiprocess systems (i.e. Onsager reciprocal relations).[47] After Onsager won the 1968 Nobel Prize his reciprocal relations became very famous, but his concept of “coupled irreversible processes” faded from people’s memory. Note that although these two concepts have some commonality they still have differences. In the next chapter, during discussion on Onsager reciprocal relations, comparison between Burk’s “coupled reactions” and Onsager’s “coupled irreversible processes” will be made. Before ending this section, a clear definition of the fundamental concept of “thermodynamic coupling” used in this book must be written out. This definition is the phenomenon for nonspontaneous process(es) and spontaneous process(es) to take place simultaneously in multi-process systems. It should be emphasized that in the same system, nonspontaneous process may take place under the driving force of other simultaneous spontaneous reaction(s), or as said by Prigogine (p. 25) [2]

> Thermodynamic coupling allows one of the reactions to progress in a direction contrary to that prescribed by its own affinity.

Here, both nonspontaneous reaction(s) and spontaneous reaction(s) are not limited to one reaction. The definition of “thermodynamic coupling” can also be written by “compensation”, as said by Clausius.(p. 364) [11]

> The second fundamental theorem, in the form which I have given to it, asserts that all transformations occurring in nature may take place in a certain direction, which I have assumed as positive, by themselves, that is, without compensation; but that in the opposite, and consequently negative direction, they can only take place in such a manner as to be compensated by simultaneous occurring positive transformations.

That is, the nonspontaneous process(es) (“negative direction”) is(are) compensated by simultaneous occurring spontaneous processes (“positive transformations”).

However, if mathematical expression is used, the definition will be simpler and clearer. Equation (4.7) could be rewritten together with using more exact symbols for entropy productions of processes, i.e. noted by subscript of p. That is

$$\mathrm{d_i}S_{\mathrm{p1}} < 0,\ \mathrm{d_i}S_{\mathrm{p2}} > 0\ \&\ \mathrm{d_i}S \geqslant 0. \tag{4.87}$$

Here, $\mathrm{d_i}S_{\mathrm{p1}}$, $\mathrm{d_i}S_{\mathrm{p2}}$ and $\mathrm{d_i}S$ are entropy productions of nonspontaneous process(es), entropy productions of spontaneous process(es) and entropy production of the whole system, respectively.

The criterion expression of thermodynamic coupling is very important in modern thermodynamics. However, rewriting Eq. (4.7) into Eq. (4.87) may not be necessary, because as noted in Eq. (4.7):

> the entropy productions due to process(es) 1 and process(es) 2 are denoted by $\mathrm{d_i}S_1$ and $\mathrm{d_i}S_2$, respectively,

has as already been clearly written. For simplicity, this book will still use the form of Eq. (4.7), [$\mathrm{d_i}S_1 < 0,\ \mathrm{d_i}S_2 > 0\ \&\ \mathrm{d_i}S \geqslant 0$].

In comparison of Eq. (4.7) with Eq. (4.5) $\mathrm{d_i}S \geqslant 0$, the general mathematical expression of the second law of thermodynamics ($\mathrm{d_i}S$ is entropy production of the whole system), it can be known that there is no restriction for entropy production of internal process. However, discussions in classical thermodynamics beginning with Carnot theorem always started with the simplest single gas system. In fact, in classical thermodynamics it was assumed that only one process or only spontaneous process(es) is included inside the system. Therefore, in classical thermodynamics entropy production of process(es) d_iS_p does always agree with entropy production of the system d_iS, i.e. $d_iS_p = d_iS$. That is, entropy production of process(es) d_iS_p had been regarded as to be satisfied $d_iS_p \geqslant 0$. For a long time that was the restriction of thermodynamics in its development. To clearly show the restriction of classical thermodynamics the cyclical-process assumption (as shown in Fig. 3.1) should be included in the mathematical expression. That is, the classical mathematical expression of the second law of thermodynamics should clearly be written in the form of Eq. (4.14), i.e. $[d_iS_p = d_iS \geqslant 0]$. Here, d_iS_p and d_iS are the entropy production of process(es) and the entropy production of system, respectively. Usually, in classical thermodynamics only single processes like those in Carnot cycle are considered. For these cases the entropy production of system is just the same as the entropy production of process. Systems with two or more spontaneous process(es) might still be considered in classical thermodynamics, if the influences between these processes were neglected. For instance, the simultaneous solutions of sugar and salt in water could still be treated by classical thermodynamics without difficulties. This is because the simultaneous solutions of sugar and salt in water has nearly no difference from the solution of sugar alone in water or the solution of salt alone in water, and the influence between these two processes might be neglected. Therefore, classical thermodynamics could firstly treat them individually, and then take the sum of these two individual entropy productions as the entropy production of the simultaneous solutions of sugar and salt in water.

Thermodynamic coupling was out of consideration in classical thermodynamics. That is the real restriction of classical thermodynamics. In other words, classical thermodynamics is only suitable for uncoupling systems. Equation (4.14) $[d_iS_p = d_iS \geqslant 0]$ is just the general mathematical expression of the second law of classical thermodynamics, while Eq. (4.7) $[d_iS_1 < 0, d_iS_2 > 0 \ \& \ d_iS \geqslant 0]$ is the general mathematical expression of the second law of modern thermodynamics. Therefore, thermodynamic coupling is the watershed between classical thermodynamics and modern thermodynamics.

As the core of modern thermodynamics, the fundamental concept was introduced directly from the second law of thermodynamics. Except that the local equilibrium approximation is necessary for any state function of nonequilibrium state, there was no necessity for any other assumptions. It is easier to understand and accept for young students. However, as shown by its meandering historical development, thermodynamic coupling had often been the key difficulty for understanding and acceptance for some classical thermodynamicists who had decades of previous working experiences and who held some out-of-date "concepts" of classical thermodynamics.

4.6 Schrödinger's "Negative Entropy" Conjecture

Erwin Schrödinger (1887–1961) was a distinguished atomic physicist. He won the 1933 Nobel Prize in physics "for the discovery of new productive forms of atomic theory". Schrödinger was very well reputed for his academic capacity. "*What is life?— The Physical Aspect of the Living Cell*" was written by Schrödinger based on his lectures delivered under the auspices of the Dublin Institute for Advanced Studies at Trinity College, Dublin, in February 1943. [48] In his book he regarded himself as "a naive physicist" "at the risk of making fools of ourselves" for discussing such complex events at that time and even up to now. He was convinced that

> The obvious inability of present-day physics and chemistry to account for such events is no reason at all for doubting that they can be accounted for by those sciences.

There is no doubt that must be correct. However, given the limitations of science at that time, an elucidation of "what is life" had obvious difficulties. Maybe in the present or in the future, with the success of "synthetic life", we will be possible to say what is life, but a full understanding of life may still remain.

In this book only thermodynamic aspects of the book "*What is Life*" will be of concern. In Schrödinger's book there was a special chapter, "Chapter 6 Order, Disorder and Entropy", on thermodynamics. Due to the historical limitations, his thermodynamic viewpoints on life basically belonged to classical thermodynamics. That is, he considered thermodynamics of life being substantially different from thermodynamics of inanimate systems. He wrote in Section, "56. Living Matter Evades the Decay to Equilibrium", that:

> What is the characteristic feature of life? When is a piece of matter said to be alive? When it goes on 'doing something', moving, exchanging material with its environment, and so forth, and that for a much longer period than we would expect an inanimate piece of matter to 'keep going' under similar circumstances.

He continued in the next Section, "57. It Feeds on 'Negative Entropy'":

> It is by avoiding the rapid decay into the inert state of "equilibrium" that an organism appears so enigmatic; so much so, that from the earliest times of human thought some special non-physical or supernatural force (vis viva, entelechy) was claimed to be operative in the organism, and in some quarters is still claimed.

Moreover, his viewpoint of "negative entropy" was:

> Thus a living organism continually increases its entropy — or, as you may say, produces positive entropy — and thus tends to approach the dangerous state of maximum entropy, which is death. It can only keep aloof from it, i.e. alive, by continually drawing from its environment negative entropy — which is something very positive as we shall immediately see. What an organism feeds upon is negative entropy. Or, to put it less paradoxically, the essential thing in metabolism is that the organism succeeds in freeing itself from all the entropy it cannot help producing while alive.

However, what is "negative entropy"? Schrödinger in Section "60. Organization Maintained by Extracting 'Order' from the Environment" said:

> How would we express in terms of the statistical theory the marvelous faculty of a living organism, by which it delays the decay into thermodynamical equilibrium (death)? We said before: 'It feeds upon negative entropy', attracting, as it were, a stream of negative entropy upon itself, to compensate the entropy increase it produces by living and thus to maintain itself on a stationary and fairly low entropy level.

Schrödinger further described:

> Hence the awkward expression 'negative entropy' can be replaced by a better one: entropy, taken

> with the negative sign, is itself a measure of order. Thus the device by which an organism maintains itself stationary at a fairly high level of orderliness (= fairly low level of entropy) really consists in continually sucking orderliness from its environment. This conclusion is less paradoxical than it appears at first sight. Rather could it be blamed for triviality. Indeed, in the case of higher animals we know the kind of orderliness they feed upon well enough, viz. the extremely well-ordered state of matter in more or less complicated organic compounds, which serve them as foodstuffs. After utilizing it they return it in a very much degraded form — not entirely degraded, however, for plants can still make use of it. (These, of course, have their most powerful supply of 'negative entropy' in the sunlight.)

Schrödinger's argumentum in the field of thermodynamics was basically not specific. The first the criterion of entropy, i.e. entropy increase principle, is merely suitable for isolated (or adiabatic) systems, but not suitable for open systems of life. Meanwhile, he perhaps did not clearly know the general concept of entropy production, i.e. positive entropy production principle, introduced through the research of P. Duhem, L. Natanson and G. Jaumann at the beginning of the 20th century. After the phenomena of thermodynamic coupling and its mathematical expression Eq. (4.7) [$\mathrm{d_i}S_1 < 0$, $\mathrm{d_i}S_2 > 0$ & $\mathrm{d_i}S \geqslant 0$] had been clearly demonstrated, could Schrödinger's "negative entropy" be regarded as "negative entropy production process" of thermodynamic coupling? I do not think so, and would like to call the "negative entropy production process" of thermodynamic coupling as non-spontaneous process. That is easier to be understood without any mysticism. Moreover, organisms (including endothermy and ectothermy creatures) may be regarded as living under isothermal isobaric conditions, so on the basis of Eq. (4.9), $\mathrm{d_i}S = -\frac{\mathrm{d}G}{T}$, the negative entropy production process of thermodynamic coupling is Gibbs free energy increase process, i.e. $(\Delta G_1)_{T,p} > 0$ in Eq. (4.13) [$(\Delta G_1)_{T,p} > 0$, $(\Delta G_2)_{T,p} < 0$ & $(\Delta G)_{T,p} \leqslant 0$]. Therefore, Schrödinger's proposition, "57. It Feeds on 'Negative entropy'", should practically change into "57. It Feeds on 'Energy'". The necessity of continuous supply of energy for maintenance of life can be elucidated. Therefore, organisms are not the perpetual machines. Phenomena of life do not violate any fundamental laws of thermodynamics. In modern thermodynamics, without mysticism, the phenomena of life are merely results of natural evolution of complex systems under the suitable continuous supply of energy.

Modern thermodynamics emphasizes evolution of life, social progression and development of science and technology. Up to now, there is no impediment for thermodynamics to understand phenomena of life. All of the difficulties in the past were only due to the limitation of classical thermodynamics. There are a lot of thermodynamic coupling in phenomena of life, but thermodynamic coupling is not merely the characteristics of life.

4.7 Chemiosmotic Coupling Theory for ATP Biosynthesis

In bioenergetics, adenosine triphosphate (ATP) plays a very important role in living bodies. [49, 50] The function of ATP is to store the available energy when food is metabolized in order to supply energy on demand to a wide variety of processes, including muscular contraction, reproduction etc. A popular view is that ATP is a special energy source that can be directly used, so that ATP is seen as a "gasoline fuel" in living bodies. Without gasoline a car cannot be driven. ATP in the human body is synthesized from adenosine diphosphate (ADP) and inorganic phosphoric acid (P_i). The synthesis of ATP from ADP is not a spontaneous process. The synthesis reaction can progress only under conditions accompanied by the input of other energy. Human cells contain hundreds or thousands of subcellular organelles called mitochondria. These organelles are approximately the size of small bacteria. Their double membrane encloses a matrix and a convoluted structure of inner membranes that contains a complex array of enzymes and proteins, as shown in Fig. 4.1. The major function of mitochondria is to convert food energy to the chemical energy of the cell (ATP). Most ATP is synthesized in mitochondria by a series of reactions that are driven by the energy released during metabolic oxidation. That is called oxidative phosphorylation. The mechanisms of oxidative phosphorylation are already rather clearly understood. The key effect of oxidative phosphorylation is the proton H^+ gradient across the inner membrane and the special structure of the inner membranes.

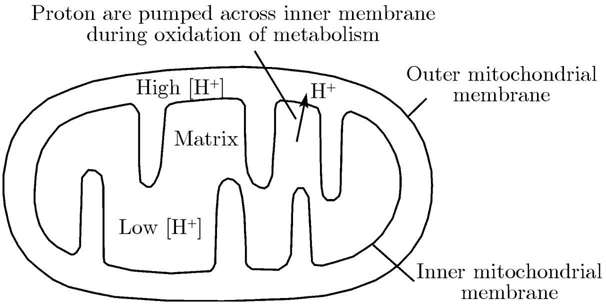


Fig. 4.1 Structure of mitochondrion and the proton (H^+) gradient across the inner membrane

A chemiosmotic coupling hypothesis can explain how the free energy generated by the electron transport chain is used to produce ATP from ADP + P_i. The members of the electron transport chain are arrayed asymmetrically in the inner mitochondrial membrane, forming loops, as shown in Fig. 4.2. The chemiosmotic coupling hypothesis states that these loops consist of alternating hydrogen atoms and electron carriers that transfer protons H^+ from the mitochondrial matrix (inside) to the cytoplasmic side (outside) of the inner mitochondrial membrane. In other words, the electron transport chain is an active proton transport system. During the oxidation, protons are pumped from the matrix side to the cytoplasmic side of the inner mitochondrial membrane. This process creates across the inner mitochondrial membrane a H^+ concentration gradient (the H^+ concentration is higher on the cytoplasmic side, i.e. outside), and an electric potential with that side positive is generated, as shown in Fig. 4.3.

The energy inherent in this gradient is sufficient to drive ATP synthesis. The free energy inherent in the concentration gradient of the proton is called the driving force of the protons. When protons reenter the mitochondrial matrix (inside of the inner mitochondrial membrane) then pass through a channel in the F_1/F_0-ATPase molecule. This results at the same time in the synthesis of ATP from ADP + P_i. Many experimental results confirm these mechanisms. For instance, once the completeness of the inner mitochondrial membrane is destroyed, ATP will not be able to synthesize. As another example, the existence of 2,4-dinitro-phenol (a reagent) allows free passage of protons through the inner mitochondrial membrane, and the concentration gradient of the protons disappears. That

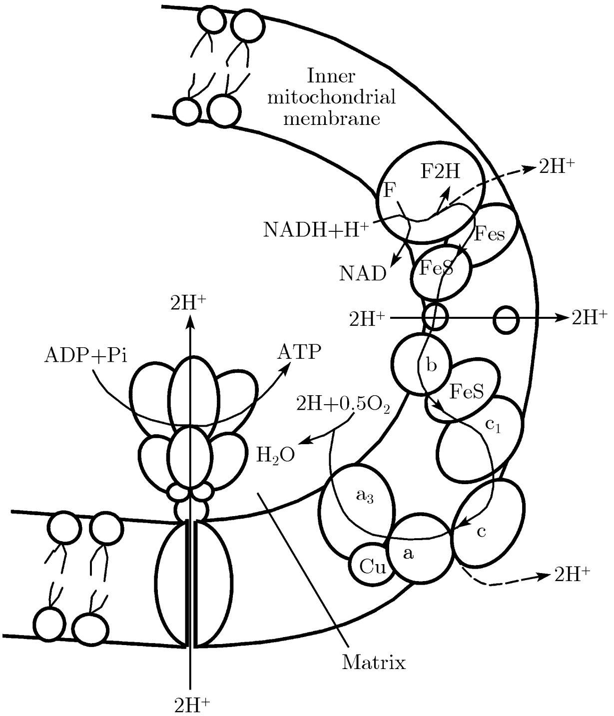


Fig. 4.2 The electron transport chain in the inner mitochondrial membrane and mechanisms of oxidative phosphorylation

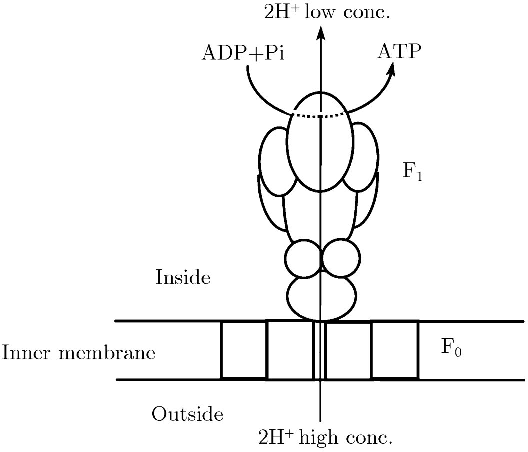


Fig. 4.3 H+ concentration gradient across the inner mitochondrial membrane drives ATP synthesis. F1/F0-ATPase is the enzyme complex that synthesizes ATP

results in the inhibition of ATP synthesis. The most powerful evidence is the ATP synthesis realized by a man-made concentration gradient, and therefore this mechanism became generally acknowledged. The chemiosmotic hypothesis was proposed by the English biochemist, Peter Mitchell, in 1961.[51] Other scientists hardly believed it at that time. Over many years, scientists searched diligently, but unsuccessfully, to find other explanations, and at last Mitchell won the Nobel Prize in 1978. The chemiosmotic hypothesis or the chemiosmotic coupling hypothesis was then renamed as the chemiosmotic theory, or the chemiosmotic coupling theory. Further detailed description of oxidative phosphorylation and other bioprocesses can be found in many books on biochemistry.

On the basis of thermodynamic coupling, the transformation from low-energetic ADP to ATP is a nonspontaneous process, $(\Delta G_1)_{T,p} > 0$; while protons (H^+) reentering the mitochondrial matrix by passing through a channel in the F_1/F_0-ATPase molecule on the inner mitochondrial membrane is a spontaneous diffusion from high concentration to low concentration, $(\Delta G_2)_{T,p} < 0$. If the whole system agrees with $[(\Delta G_1)_{T,p} > 0,\ (\Delta G_2)_{T,p} < 0\ \&\ (\Delta G)_{T,p} \leqslant 0]$, then the biosynthesis of ATP can take place successful. Therefore, the biosynthesis of ATP not only agrees completely with the second law of thermodynamics, but also is a powerful qualitative demonstration for the criterion of thermodynamic coupling. Even after Mitchell won the Nobel Prize in 1978, the chemiosmotic coupling theory has rarely been included in thermodynamics textbooks in the physics and chemistry disciplines, although it has been an important feature in biochemistry textbooks.

4.8 Classical and Traditional Classifications of Thermodynamics

4.8.1 Classical Classifications of Thermodynamics

Since the setup of Clausius inequality $\mathrm{d}S - \frac{\delta Q}{T} \geqslant 0$ or the entropy increase principle $(\mathrm{d}S)_{\mathrm{iso}} \geqslant 0$ for the second law of thermodynamics, the mathematical expression had already been shown that classical thermodynamics consists of an equality part and an inequality part. Obviously, on the establishment of thermodynamics there was no question about what is "classical thermodynamics" or "modern thermodynamics". At the beginning of the 20th century, especially during the 1920s–1930s, the term "classical thermodynamics" emerged, but after then for a long time the term "modern thermodynamics" was not generally recognized nearly in the whole 20th century. Classical thermodynamics was usually regarded as the whole discipline of thermodynamics. Therefore, the equality part of the mathematical expression of the second law [i.e. the equality part of Clausius inequality, the entropy increase principle or Gibbs free energy form for isothermal isobaric conditions, $(\mathrm{d}G)_{T,p} = 0$] was regarded as the criterion (the necessary and sufficient conditions) of a system in its equilibrium state, while the inequality was regarded as the criterion of a system in its nonequilibrium state. That is, thermodynamics was divided into equilibrium thermodynamics and nonequilibrium thermodynamics. That is the classical classification of thermodynamics, as shown in Fig. 4.4. The criterion in square brackets for each field in Fig. 4.4 could also be written in Gibbs free energy form for isothermal isobaric systems.

The combination of the equality part of the second law (i.e. $T\mathrm{d}S - \delta Q = 0$, and the equality of the first law, such as $\mathrm{d}U = \delta Q - \delta W$) for closed systems easily produced a huge and rather rigorous academic field of classical equilibrium thermodynamics. The nonequilibrium thermodynamics of classical thermodynamics was considered to be "drowned" by the achievement of equilibrium thermodynamics. However, the existence of the inequality part of the second law, i.e. $T\mathrm{d}S - \delta Q > 0$, was never denied. That might be due to a lot of difficulties arising from the inequality form in developing nonequilibrium thermodynamics.

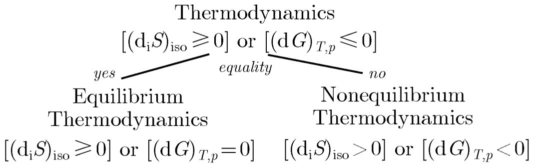


Fig. 4.4 Classical Classifications of Thermodynamics

4.8.2 Traditional Classifications of Thermodynamics

Due to the effort over about a hundred years, in the middle of the 20th century much progress in nonequilibrium thermodynamics had been achieved. Therefore, nonequilibrium thermodynamics had been subdivided into linear nonequilibrium thermodynamics and nonlinear nonequilibrium thermodynamics. The linear and nonlinear relations between thermodynamic forces and thermodynamic flows of the nonequilibrium system were taken as the criteria of linear nonequilibrium thermodynamics and nonlinear nonequilibrium thermodynamics, respectively. In this way, at the end of the 20th century the whole field of thermodynamics had been divided into three branches of thermodynamics. These were equilibrium thermodynamics, linear nonequilibrium thermodynamics and nonlinear nonequilibrium thermodynamics.[2, 3]. Sometimes, these three fields were called thermodynamics for reversible processes, thermodynamics for linear irreversible processes and thermodynamics for nonlinear irreversible processes, respectively.

4.8.2.1 Equilibrium thermodynamics (or thermodynamics for reversible processes)

There is no irreversible process in the corresponding systems. That is to say, both thermodynamic force and thermodynamic flow are equal to zero in the system. This part in classical thermodynamics is called classical equilibrium thermodynamics. Classical equilibrium thermodynamics had been well developed, has played a very important role in physics, chemistry, engineering science and industrial manufacturing, and will continue to play an important role in the future. However, classical equilibrium thermodynamics is not suitable for complex systems including thermodynamic coupling, such as many living systems and systems with new technology, and thus it is mainly used for simple isolated and closed systems without thermodynamic coupling. For an open system, both entropy change of the system $\mathrm{d}S_{\mathrm{syst}}$ and entropy change of its surroundings $\mathrm{d}S_{\mathrm{surr}}$ should be considered, i.e. $\mathrm{d}S_{\mathrm{iso}} = (\mathrm{d}S_{\mathrm{syst}} + \mathrm{d}S_{\mathrm{surr}})$, and it is still treated as an isolated system. In fact, $\mathrm{d}S_{\mathrm{surr}}$ is more difficult to calculate.

4.8.2.2 Linear nonequilibrium thermodynamics (or thermodynamics for linear irreversible processes)

The character of nonequilibrium thermodynamics was such that it was regarded as a way of studying systems comprising irreversible processes. Due to the irreversible process taking place in the system the thermodynamic force X and the thermodynamic flow J were not equal to zero. The thermodynamic force X is the cause of the thermodynamic flow J, so it can be considered that there is a functional relation between them, $J = F(X)$. Supposing that the functional relation is continuous and it is possible to express it as a Taylor function at the reference point, in which the force and the flow both equal zero, then for a simple irreversible process

$$J = J(X) = J_0(X_0) + \left(\frac{\partial J}{\partial X}\right)_0 (X - X_0) + \frac{1}{2}\left(\frac{\partial^2 J}{\partial X^2}\right)(X - X_0)^2 + \cdots. \tag{4.88}$$

At the reference point, $X_0 = 0$, $J_0 = 0$, so

$$J = LX + \frac{1}{2}\left(\frac{\partial^2 J}{\partial X^2}\right)_0 X^2 + \cdots. \tag{4.89}$$

Here, $L = \left(\frac{\partial J}{\partial X}\right)_0$. When the deviation from the reference point is so small that the higher-power terms of X are much smaller and can be neglected, then

$$J = LX. \tag{4.90}$$

From experimental observations, such a form of relationship has in fact been obtained. For instance, the magnitude of the heat flow is usually directly proportional to the temperature gradient $\left(\frac{\partial T}{\partial r}\right)$. The magnitude of the diffusion flow is also directly proportional to the concentration gradient. These are Fourier law for heat conduction and Fick law for diffusion. These macroscopic laws coming from the summarization of experimental phenomena are called phenomenological laws. The proportional coefficients L are called phenomenological coefficients. If L is basically independent of X (i.e. L is approximately a constant with respect to the change of X), then such an irreversible process can be called an irreversible process in its linear regime. The corresponding field for thermodynamic research was called linear nonequilibrium thermodynamics, or thermodynamics for linear irreversible processes.

4.8.2.3 Nonlinear nonequilibrium thermodynamics (or thermodynamics for nonlinear irreversible processes)

When the thermodynamic force is not too small in the system, the relation formula will not be linear between the thermodynamic force J and the thermodynamic flow X of the irreversible process. Under such situations the higher-power terms in the relation formula between the thermodynamic flow J and the thermodynamic force X must be considered. According to the research done by Prigogine et al., some macroscopically ordered phenomena may appear in nonlinear nonequilibrium systems. Prigogine called them self-organization phenomena. Some other scientists called them synergetical phenomena. The appearance and maintenance of macroscopic phenomena of ordered self-organization must depend on dissipation of energy (including the input of external energy and/or internal energy). Therefore, Prigogine used the term dissipative structures to refer to these special nonequilibrium macroscopic ordered states of the systems.

Based on the above-mentioned three fields, and especially due to the fact that Onsager and Prigogine won Nobel Prizes in 1968 and 1977, respectively, the traditional classification of thermodynamics (as shown in Fig. 4.5) had also been regarded as a classification of "modern thermodynamics" at that time.

The traditional classification of thermodynamics, as shown in Fig. 4.5, was proposed during the 1970s – 1980s, so both the classical expressions of the second law of thermodynamics, i.e. entropy increase principle $[(\mathrm{d}S)_{\mathrm{iso}} \geqslant 0]$, and the general expression of the second law of thermodynamics, i.e. positive entropy production principle $\mathrm{d_i}S \geqslant 0$, are used. The substitution of $\mathrm{d_i}S \geqslant 0$ for $[(\mathrm{d}S)_{\mathrm{iso}} \geqslant 0]$ did not bring obvious changes in the traditional classification of thermodynamics, so some scientists might consider that the change from entropy increase principle into positive entropy production principle might be only a change in form.

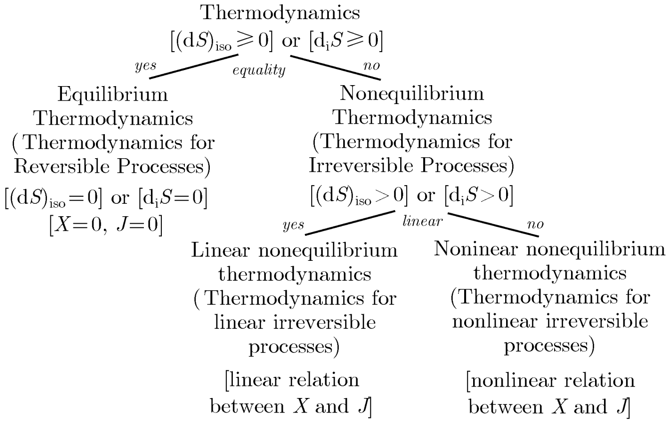


Fig. 4.5 Traditional classification of thermodynamics

Sometimes, for unifying the criteria in classification and "avoiding dividing the subject into equilibrium and nonequilibrium thermodynamics", the relationship between thermodynamic force and thermodynamic flow was taken as the criteria of the first-order classification:

1. Both thermodynamic force and thermodynamic flow equal zero $[X = 0\,\&\,J = 0]$;
2. Both thermodynamic force and thermodynamic flow do not equal zero, and there is a linear relation between X and J;
3. Both thermodynamic force and thermodynamic flow do not equal zero, and there is a nonlinear relation between X and J.

In this way, thermodynamics was directly divided into three subfields: equilibrium thermodynamics, linear nonequilibrium thermodynamics and nonlinear nonequilibrium thermodynamics. In fact, such a classification of thermodynamics was a reflection of the practical path in the development of thermodynamics. Therefore, these three subfields were also known as three development stages of thermodynamics. All of them were regarded as "modern thermodynamics" by D. Kondepudi and I. Prigogine at the end of the 20th century. However, such a "modern thermodynamics" was not generally recognized. It was mentioned that [4]

> most texts that are currently available to students still treat thermodynamics very much as it was presented in the 19th century.

Besides, the research objects in classical thermodynamics are simple uncoupling systems, so in textbooks of classical thermodynamics the equality of the second law was always called equilibrium thermodynamics. Thus, linear nonequilibrium thermodynamics was often called close-to-equilibrium nonequilibrium thermodynamics, or near-equilibrium nonequilibrium thermodynamics, while nonlinear nonequilibrium thermodynamics was often called far-from-equilibrium nonequilibrium thermodynamics. These appellations may also be found in Onsager's and Prigogine's papers. In discussions in this book later it will be clear that linear nonequilibrium thermodynamics and close-to-equilibrium nonequilibrium thermodynamics are different concepts, and they should not be treated as synonyms.

In about 1970, a new technology in the activated low-pressure diamond growth was achieved first in the former Soviet Union, but such a new technology suffered from suspicion and ridicule for 16 years. At last, in 1986 this technology was worldwidely recognized. However, in thermodynamics it was still regarded as "violating the second law of thermodynamics". In fact, that was a good opportunity to confirm thermodynamic coupling, i.e. $[\mathrm{d_i}S_1 < 0,\ \mathrm{d_i}S_2 > 0\ \&\ \mathrm{d_i}S \geqslant 0]$. After the phenomenon of thermodynamic coupling and the criterion of thermodynamic coupling had been quantitatively demonstrated, the limitation of classical thermodynamics was gradually discovered, and classification of thermodynamics had to be drastically changed, as is mentioned in the next section. [6~9,26~46]

4.9 Modern Classification of Thermodynamics

Thermodynamic coupling is the watershed between classical thermodynamics and modern thermodynamics. Therefore, based on systems including or not including thermodynamic coupling, the whole discipline of thermodynamics could immediately be divided into two first-order subfields: classical thermodynamics and modern thermodynamics, as shown in Fig. 4.6.

From the first-order division it can be found that: The criterion of thermodynamic coupling $[\mathrm{d_i}S_1 < 0,\ \mathrm{d_i}S_2 > 0\ \&\ \mathrm{d_i}S \geqslant 0]$ is just the mathematical expression of the second law of thermodynamics for complex multi-process systems on the basis of local equilibrium approximation. This expression can also be called the mathematical expression of the second law of modern thermodynamics. For simple uncoupling system, the expression of the second law of thermodynamics should be clearly written as Eq. (4.14)

---

[4] http://as.wiley.com/WileyCDA/WileyTitle/productCd0471973947.html. Cited 2004–2008

$[d_iS_p = d_iS \geqslant 0]$. Nonspontaneous processes were not considered in classical thermodynamics because classical thermodynamics was only used for uncoupling systems. There is no limitation for application to equilibrium systems in classical thermodynamics, but the application to nonequilibrium systems is very limited in classical thermodynamics. In most thermodynamics textbooks the simple connection between the equality of the expression of the second law and equilibrium system or reversible process was unsuitable. For simple systems that is still OK, but in general for any macroscopic systems that may be incorrect. In other words, there had been a bug hidden in classical or traditional concepts of thermodynamics for a long time, because the equality part of the mathematical expression of the second law of thermodynamics for multi-process systems only means nondissipation and the corresponding system may be not in equilibrium.

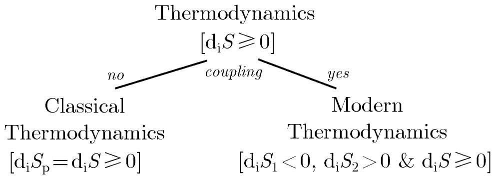


Fig. 4.6 Thermodynamic coupling is a watershed between classical thermodynamics and modern thermodynamics

The second-order division of thermodynamics is based on equality or inequality. Now the classical thermodynamics and modern thermodynamics will be divided into two parts, individually. The equality part of classical thermodynamics can be called equilibrium thermodynamics (the same as its original name); while the inequality part of classical thermodynamics can be called nonequilibrium thermodynamics (also the same as its original name), as shown in Fig. 4.7. It is not necessary to change these names into "classical nondissipative thermodynamics" and "classical dissipative thermodynamics" for unification of the nomenclature, because both names of equilibrium thermodynamics and nonequilibrium thermodynamics are well-accepted.

In comparison with Fig. 4.4, it can be found that within the applicability of classical thermodynamics both Fig. 4.7 and Fig. 4.4 are nearly the same, but Fig. 4.7 is clearly written only for classical thermodynamics, and Fig. 4.4 was incorrect to use the mathematical expression for the whole of thermodynamics. Now it is very clear that classical thermodynamics is a part of thermodynamics for uncoupling systems. In other words, classical thermodynamics was deduced strictly on the basis of Clausius assumptions and his mathematical expression of the second law of thermodynamics together with the mathematical expression of the first law of thermodynamics. Classical thermodynamics is rigorous inside its applicability framework, but the basic cyclical-process assumption must be remem-

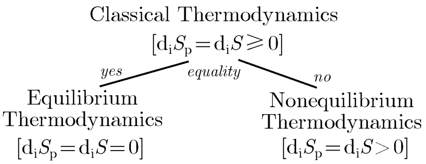


Fig. 4.7 Classification of Classical Thermodynamics

bered. The basic cyclical-process assumption was included in Carnot theorem and introduced by Clausius during deduction of the classical mathematical expression of the second law of thermodynamics. That decided the severe limitation of classical thermodynamics for studying irreversible processes. In other words, it was just the strictness of classical thermodynamics that decided the restriction of its usage for a continuous going irreversible process, because that does not agree with the cyclical-process assumption.

The equality part of the mathematical expression of the second law of thermodynamics for coupling systems (i.e. mathematical expression of the second law of modern thermodynamics) may be called nonequilibrium nondissipative thermodynamics, modern nondissipative thermodynamics, or simply called nondissipative thermodynamics, which is a new field in modern thermodynamics or in thermodynamics. Therefore, the inequality part of the mathematical expression of the second law of thermodynamics for coupling systems should be called dissipative thermodynamics or modern dissipative thermodynamics, but should not be called nonequilibrium thermodynamics in order to prevent confusion with nonequilibrium thermodynamics in classical thermodynamics. Both Onsager reciprocal relations and Prigogine dissipative structures would belong to dissipative thermodynamics. A third-order division could be made in dissipative thermodynamics. The third-order division may be based on linearity between thermodynamic force X and thermodynamic flow J of irreversible process, i.e. [linear relation between X and J] or [nonlinear relation between X and J]. In this way, dissipative thermodynamics is divided into linear dissipative thermodynamics and nonlinear dissipative thermodynamics. Of course, the third-order division may also be based on the value of entropy production, i.e. [linear relation, if $\mathrm{d_i}S$ (a little)> 0] or [nonlinear relation, if $\mathrm{d_i}S >> 0$], as shown in Fig. 4.8, but the third-order division is not so substantial as in the first-order and in the second-order divisions.

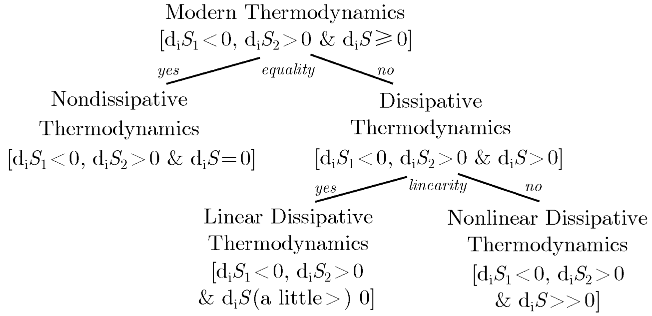


Fig. 4.8 Classification of modern thermodynamics

If we put Fig. 4.6, Fig. 4.7 and Fig. 4.8 together, then a basic and complete modern classification of thermodynamics can be obtained, as shown in Fig. 4.9.

Of course, this is the most complete classification of the thermodynamics discipline up to now. It also clearly shows an overview of thermodynamics in the current 21^{st} century. Such a classification is only based on a deeper understanding the second law of thermodynamics in the current 21^{st} century. The complete classification is not unique. If the criteria of the first-order division and the second-order division in Fig. 4.9 are exchanged with each other, then another complete modern classification of thermodynamics can be obtained. In such a way, the names of some branch fields should undergo some small modification. For instance, after the first-order division two subfields couuld be directly called "nondissipative thermodynamics" and "dissipative thermodynamics", while after the second-order division four subfields could be renamed "classical nondissipative thermodynamics" (i.e. equilibrium thermodynamics), "modern nondissipative thermodynamics", "classical dissipative thermodynamics"(i.e. nonequilibrium thermodynamics) and "modern dissipative thermodynamics". Such changes are quite different from current convection, and this could cause some confusion. Therefore, in this book only Fig. 4.9 is recommended and this is called the basic and complete modern classification of thermodynamics.

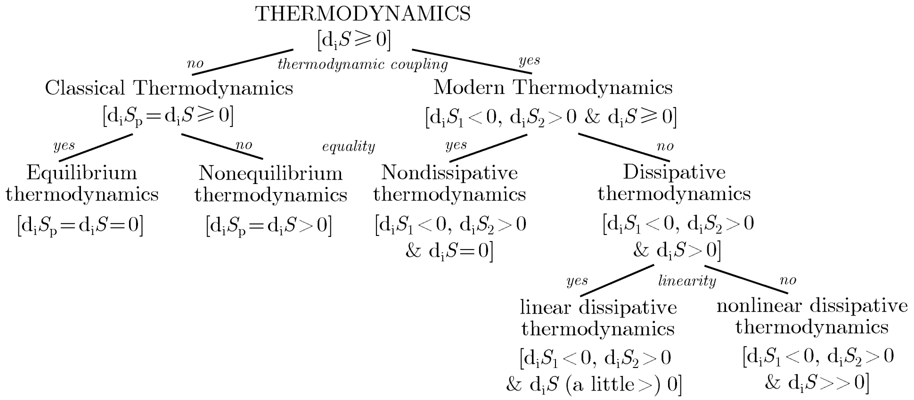


Fig. 4.9 The basic and complete modern classification of thermodynamics, or called the second law of thermodynamics of the current 21$^{\mathrm{st}}$ century

4.10 Extended Carnot Theorem

Carnot theorem was the foundation stone of thermodynamics and classical thermodynamics, and it was the starting point for the second law of thermodynamics. The extended Carnot theorem is the extended foundation stone of thermodynamics and modern thermodynamics and can be regarded as a new starting point of the second law of modern thermodynamics. In general, the "reversibility" is not a necessary condition for the highest conversion efficiency of energy, but "nondissipation" is really the necessary and sufficient condition for the highest conversion efficiency of energy. Correspondingly, there is a new field of nondissipative thermodynamics together with the field of dissipative thermodynamics.

It is well known that the basic laws of thermodynamics were established on the basis of a large number of daily human macroscopic experiences, but not deduced from any more basic law of other disciplines of science. Carnot theorem told us: the reversible heat engine is of the highest conversion efficiency of energy among all other heat engines working between the same hotter heat reservoir and the same colder heat reservoir. On the basis of Carnot theorem, Kelvin and Clausius proposed the written statements of the second law of thermodynamics, and then Clausius introduced the entropy function and setup the quantitative mathematical expressions of entropy increase principle and Clausius inequality. Till now, the mathematical expressions of the second law of classical thermodynamics can be expressed as Eq. (4.14) [$\mathrm{d_i}S_\mathrm{p} = \mathrm{d_i}S \geqslant 0$]. Now backwardly, from Eq. (4.14) it can be known that just the equality part, i.e. [$\mathrm{d_i}S_\mathrm{p} = \mathrm{d_i}S = 0$], corresponds to the highest conversion efficiency of heat engine. For simple systems, such as heat engines, entropy production of only reversible process equals zero, so its conversion efficiency must be the highest. That was the important experience of human beings at that time (1824) as summarized by Carnot.

At the same time we know that the nature of the second law of thermodynamics is the irreversibility of time for the development of macroscopic systems, i.e. the characteristics of "arrow of time". However, Carnot theorem did just emphasize "reversibility". Such a paradox existed always in classical thermodynamics, and that was just the root of limitation in classical thermodynamics. Such a status might have lasted more than 180 years. Such a status had rarely been found in modernization of other disciplines of science, or it could be said, up to now this is unique. The human experience has developed, and the framework and the deepness of applicability of thermodynamics are also in development. The research objects of thermodynamics have been extended from heat-work conversion into any kind of energy-conversion under thermodynamic consideration. That is, the extended Carnot theorem is needed for summarization of new experiences of human beings at the present time.

As said above, thermodynamic coupling is the watershed between modern thermodynamics and classical thermodynamics, and it is also the core of modern thermodynamics, so the connotation of the second law of thermodynamics is very abundant. From the mathematical expression of thermodynamic coupling, Eq. (4.7) [$\mathrm{d_i}S_1 < 0$, $\mathrm{d_i}S_2 > 0$ & $\mathrm{d_i}S \geqslant 0$], the equality part means that nondissipative process is the necessary and sufficient condition for the highest conversion efficiency of energy, or we can say, that is the extended Carnot theorem.

The full name of thermodynamic coupling is thermodynamic coupling between simultaneous irreversible processes in the same macroscopic system. In particular, thermodynamic coupling between nonspontaneous process and spontaneous process may drive nonspontaneous process to take place, while the nonspontaneous is originally impossible to take place alone. The most direct application of the extended Carnot theorem was the establishment of a new field of nondissipative thermodynamics. Carnot reversible heat engine is an idealized limitation. We may approach the limitation, but never completely reach it. Macroscopic nondissipative conversion of energy is also an idealized limitation of science. We may approach the limitation, but never completely reach it either. That never lowers the important meanings of scientific limitation.

For instance, the chemical energy of the reaction between methane (CH_4) and oxygen (O_2) in fuel-battery could be converted into electric energy. Thus should we pursue a kind of reversible fuel-battery for getting the highest conversion efficiency of energy? Obviously, from carbon dioxide (CO_2) and water (H_2O) produced and the electric energy released by the fuel-battery, the equal amount of methane and oxygen consumed are impossible to be obtained because during the synthesis of methane a series of organic hydrocarbon, such as ethane, ethylene, acetylene etc., will inevitably be obtained. Meanwhile, it can be imagined that if the chemical energy can completely convert into the electric energy without dissipation, then its efficiency must be the highest. The "reversibility" in Carnot theorem is not a general necessary condition, but "nondissipative process" is just the necessary and sufficient condition for the highest macroscopic conversion efficiency of energy. Therefore, the extended Carnot theorem is that nondissipative process is of the highest macroscopic conversion efficiency of energy. The original Carnot theorem is a special example of the extended Carnot theorem for cyclical processes.

The extended Carnot theorem is the foundation stone of modern thermodynamics, but it was proposed a long time after the proposal of the concept of thermodynamic coupling, which was fully confirmed by experiments of the activated low-pressure diamond growth.

4.11 Dissipation (or Entropy Production) Decrease Theorem

From the mathematical expression of thermodynamic coupling [$\mathrm{d_i}S_1 < 0, \mathrm{d_i}S_2 > 0 \,\&\, \mathrm{d_i}S \geqslant 0$] it can be found that spontaneous process (i.e. positive entropy production process or dissipative process, $\mathrm{d_i}S_2 > 0$) is a kind of potential driving force to drive any possible nonspontaneous process (i.e. negative entropy production process or dissipative decrease process here, $\mathrm{d_i}S_1 < 0$), if the entropy production of the system agrees with $\mathrm{d_i}S \geqslant 0$. For simple systems, there is no possibility for nonspontaneous process to take place.

As mentioned in Chapter 1, from the mathematical expression of thermodynamic coupling, i.e. [$\mathrm{d_i}S_1 < 0$, $\mathrm{d_i}S_2 > 0$ & $\mathrm{d_i}S \geqslant 0$], another natural rule, dissipation (or entropy production) decrease theorem, can be obtained. Under the conditions of continuous energy supply from exterior, the available energy of nonequilibrium systems does not always tend to the maximum of dissipation or to maximum of entropy production. That is, in coupling systems spontaneous irreversible process does always incline to utilize any useful energy in the form of transformation or storage into nonspontaneous process, i.e. negative entropy production process. The framework of applicability of this theorem is just the same as that of modern thermodynamics, including nondissipative thermodynamics, and linear and nonlinear dissipative thermodynamics for coupling systems. We can also say that this theorem is a generally applicable theorem, but in simple systems the corresponding nonspontaneous process could not form due to the systems being too simple. For instance, in a system composed of A and B gas mixture, heat conduction may cause a nonspontaneous diffusion against the concentration gradient; but in a gas of single molecules, heat conduction could not form a nonspontaneous diffusion. This example of thermodiffusion might be the simplest coupling system, and it is only concerned with physics. More examples of thermodynamic coupling, such as that discussed in Section 4.7 Chemiosmotic Coupling Theory for ATP Biosynthesis above, will be discussed throughout this book. In particular, a special topic concerned with the relationship between complexity of systems and thermodynamic coupling will be discussed in Section 9.5 Thermodynamic Weakness of Physicists — Complex Systems of the last chapter. In fact, this theorem is the thermodynamic basis and guidance for biological evolution in nature. The conversion efficiency of energy in living bodies is often very high, and that is the result due to the continuous complication and the continuous evolution. Therefore, this theorem is of deep academic significance, and a basic principle which should be followed in social progression, and in the development of science and technology.

Based on the mathematic expressions of thermodynamic coupling, i.e. [$\mathrm{d_i}S_1 < 0, \mathrm{d_i}S_2 > 0 \,\&\, \mathrm{d_i}S \geqslant 0$], it is sure that $\mathrm{d_i}S = \mathrm{d_i}S_1 + \mathrm{d_i}S_2 = |\mathrm{d_i}S_2| - |\mathrm{d_i}S_1| \geqslant 0$. If inside the system only spontaneous process 2 takes place, entropy production of the system will be $\mathrm{d_i}S = \mathrm{d_i}S_2 \geqslant 0$.

Due to thermodynamic coupling, entropy production of the system will be smaller, and equals $|d_iS_2| - |d_iS_1| \geqslant 0$. That is the reason for the name of dissipation (or entropy production) decrease theorem. At the same time, the system is still agree with the positive entropy production principle, i.e. $[d_iS \geqslant 0]$. It should also be emphasized that dissipation (or entropy production) decrease theorem is completely different from the entropy production minimization principle, which was proposed by Prigogine and will be discussed in Section 5.4 of the next chapter.

References

1. Kondepudi D, Prigogine I. Modern Thermodynamics — From Heat Engines to Dissipative Structures [M]. New York: John Wiley & Sons, 1998. http://www.wiley.com/.
2. Prigogine I. Introduction to Thermodynamics of Irreversible Processes [M]. 3rd edn. New York: Interscience Publishers, John Wiley & Sons, 1967.
3. Li R-S. Nonequilibrium Thermodynamics and Dissipative Structure [M], Beijing: Tsinghua University Press, 1986.
4. Hillert M. Phase Equilibria, Phase Diagrams and Phase Transformations Their Thermodynamic Basis [M]. Boston: Cambridge Univ Press, 1998.
5. Fu X-C, Shen W-X, Yao T-Y. Physical Chemistry [M]. 4th edn. Beijing: High Education Press, 1990, 172.
6. Wang J-T, Zhang D W, Liu Z-J. Thermodynamic Coupling Model for Low Pressure Diamond Growth from the Vapor Phase [M]. Beijing: Science Press, 1998, reprinted in 2000, 175 pages.
7. Wang J-T. Phase Diagrams of Stationary Nonequilibrium States — Thermodynamics for Low Pressure Diamond Growth from the Vapor Phase [M]. Beijing: Science Press, 2000, 212 pages.
8. Wang J-T. Nonequilibrium Nondissipative Thermodynamics – With Application to Low-Pressure Diamond Synthesis [M]. Heidelberg: Springer, 2002.
9. Wang J-T. Modern thermodynamics – and a whole view of thermodynamics [M]. Shanghai: Fudan Univ. Press, 2005, 223 pages
10. Lin Z-H (ed.). Thermodynamics and Statistic Physics [M]. Beijing: Beijing University Press, 2007, 1-2.
11. Clausius R. Mechanical Theory of Heat [M]. London: John van Voorst, 1867, Ninth memoir, 364.
12. Burk D. J Phys Chem [J], 1931, 35: 432.
13. Glasstone S. Textbook of Physical Chemistry [M]. 2nd edn 7th printing. England: MacMillan, 1951, 230.
14. De Donder Th, Van Rysselberghe P. Thermodynamic Theory of Affinity - A book of principles [M]. 2nd edn. Stanford: Stanford University Press, 1936.
15. Van Rysselberghe P. J Phys Chem [J], 1937, 41: 787.
16. Van Rysselberghe P. Bull Ac Roy Belg (Cl Sc) [J], 1936, 22: 1330; 1937, 23: 416. Presente par De Donder M.
17. Zheng Z-Z. Introduction to Thermodynamics for Irreversible Processes and Modern Reaction Kinetics [M]. Beijing: High Education Press, 1983, 50-51.
18. Prigogine I, Defay R. Chemical Thermodynamics [M]. translated by Everett DH. London: Longmans Green and Co Inc, 1954, 38.

19. Southerland W M. Biochemistry [M]. New York: Churchill Livingstone, 1990, 53-55.
20. Castellan G W. Physical Chemistry [M]. 3rd edn. London: Addison-Wesley Pub Co, 1983, 246.
21. Leidler K J, Meiser J H. Physical Chemistry [M] Benjamin/Cummings Publishing Inc, 1982, 154.
22. Boudart M. J Phys Chem [J], 1982, 87: 2766-2789.
23. Boudart M. In: Catalyst Design [M]. New York: Wiley Interscience Pub, Chapter 5, 1987.
24. Li R-S. Catalysts and thermodynamics coupling of chemical reactions. Acta Chemica Sinica [J], Eng edn, 1989(4): 305-310.
25. Koltz I M, Rosenberg R M. Chemical Thermodynamics [M]. New York: Benjamin, 1972, 150.
26. Wang J-T. The Second Law of Thermodynamics in the Current 21st Century (plenary lecture). 2007 National Symp on Thermodynamics and Statistics [C], Yanji, China: 6 Aug 2007.
27. Wang J-T, Zheng P-J. The Second Law in Modern Thermodynamics. 19th Internat Conf on Chem Thermodynamics [C]. Boulder, CO, USA: 30 Jul - 4 Aug 2006, http://www.symp16.nist.gov/pdf/p2086.pdf. Cited 20 Apr 2008.
28. Wang J-T, Zheng P-J. Thermodynamics for Carat-size Synthetic Diamonds. 19th Internat Conf on Chem Thermodynamics [C]. Boulder, CO, USA: 30 Jul - 4 Aug 2006. http://www.symp16.nist.gov/pdf/p2082.pdf. Cited 20 Apr 2008.
29. Wang J-T, Zheng P-J. Extended Carnot Theorem. 19th Internat Conf on Chem Thermodynamics [C]. Boulder, CO, USA: 30 Jul - 4 Aug 2006. http://www.symp16.nist.gov/pdf/p745.pdf. Cited 20 Apr 2008.
30. Wang J-T. Modern Thermodynamics in the 21st Century (plenary lecture). 2006 National Symp on Thermodynamics and Statistics [C]. Lanzhou, China: 17 Jul 2006.
31. Wang J-T. Modern Thermodynamics of the 21st Century and Nonequilibrium Phase Diagrams for Carat-size Synthetic Diamond. Proc. of the 13th National Symp. (Chinese and Japanese Join Meeting) on Phase Diagrams [C]. Xiamen, China: 9 Nov 2006, 245.
32. Wang J-T. A New Field of Nonequilibrium Nondissipative Thermodynamics (plenary lecture). Abstracts of Russian International Conf on Chemical Thermodynamics (RCCT2005) [C]. Moscow, Russia: 31 Jun - 2 Jul, 2005.
33. Wang J-T, Zhang D W, Yu W-F. The Second Law of Thermodynamics for Advanced Thin Films. 5th Internat Con. on Thin Film Physics and Applications [C]. Shanghai: 31 May - 2 Jun 2004.
34. Wang J-T, Zhang D W, Yu W-F. Modern Thermodynamics for CVD Diamond. 9th Internat Symp on Diamond Materials [C]. Extended Abstract of The Electrochemical Society. Honolulu, Hawaii: 8 Oct 2004.
35. Wang J-T. Modern Physics Lett B [J], 2002, 16(23 & 24): 885-888.
36. Wang J-T, Liu Z-K, Zhang D W. In: Frontiers of Solid State Chemistry [M]. Proc of the International Symposium on Solid State Chemistry in China, Eds. By Feng S-H and Chen JS. Singapore: World Scientific, 2002, 541-548.
37. Wang J-T. Modern thermodynamics for activated chemical vapor deposition (CVD) processes. In: Chemical Vapor Deposition, (extended abstract) 2nd Asian Conference, at Gyeongju, Korea, 28-30 May 2001 [C]. Korea: Ponhang University of Science and Technology, 2001, 63-66.
38. Wang J-T. Nonequilibrium nondissipative thermodynamics — and its application to diamond film deposition. In: Solid-State and Integrated-Circuit Technology, 6th International Conf, at Shanghai, China, 22-25 Oct 2001 [C]. Beijing: IEEE Press & People's Posts and Telecommunication Publishing House, 2001 vol 2: 1421-1426.

39. Wang J-T. Modern thermodynamics in CVD of hard materials. In: The Science of Hard Materials, 7th International Conference [C]. Ixtapa, Mexico: 5-9 Mar 2001, 105-106; Wang J-T, Shen J-Y, Zhang D W. J Refractory Metals and Hard Materials [J], 2001, 19(4-6): 461-466.
40. Wang J-T. Modern thermodynamics for phase diagrams of stationary states. In: Phase Diagram Calculation (CALPHAD XXIX) (Abstracts). 29th International Conference [C]. MIT, Cambridge, Massachusetts, USA: 18–23 Jun 2000, 8.
41. Wang J-T. Systematization of modern thermodynamics. In: Chemical Thermodynamics (ICCT-2000), abstracts of 16th IUPAC International Conference [C], Halifax, Canada: 6-11 Aug 2000, 150.
42. Wang J-T. Modern thermodynamics for advanced materials research. Advanced Materials. In: Internat Conf on Eng & Tech Science 2000, vol 1, Session 3, at Beijing, China: Oct 2000, ed by Song J, Yin R-Y, [C]. Beijing: New World Press, 2000, 864-870.
43. Wang J-T. Physics (Beijing) [J], 2003, 32(1): 9-15.
44. Wang J-T, Zhang D W. Nonequilibrium Nondissipative Thermodynamics & Calculation of Nonequilibrium Phase Diagrams. In: Proc. of 11th National Symp. on Phase Diagrams [C]. Xining: 20–24 Aug 2002, 131-134; J Salt Lake Research [J], 2003, 11(1): 62-65.
45. Wang J-T. University Chemistry [J], 2002, 17(2): 29-34.
46. Wang J-T. Physics (Beijing) [J], 2000, 29(9): 524-530.
47. Onsager L. Physical Review [J], 1931, 37: 405; 38: 2265.
48. Schrödinger E. What is life [M]. 1st edn. London: Cambridge University Press, 1944; Doubleday, 1956, 71.
49. Shen R-Q, Gu Q-M. Textbook of Biochemistry [M]. Beijing; High Education Press, 1993, 328.
50. Zubay G. Biochemistry [M]. Massacusetts: Addison-Wesley Publishing Company, 1983, 387-395.
51. Mitchell P. Science [J], 1979, 206: 1148-1159.

Chapter 5
Dissipative Thermodynamics

Abstract Thermodynamics is a discipline of science for research on development and evolution of macroscopic systems. Carnot theorem emphasized reversibility of process, so "reversible" and "irreversible" processes, and related "equilibrium" and "nonequilibrium" systems, and "quasistatic" process etc. are the most fundamental concepts of classical thermodynamics. However, in the first hundred years of thermodynamics development, no sufficient attention was paid to another fundamental conceptual pairing of "nondissipation" and "dissipation". For simple systems, "nondissipative" might equal "reversible", while "dissipative" might equal "irreversible". Therefore, in classical thermodynamics it was really not necessary to introduce the pair of fundamental concepts, "nondissipative" and "dissipative", but they are very important in modern thermodynamics which deals with complex multiprocess systems. In modern thermodynamics, the starting point would be the extended Carnot theorem: "nondissipative process" is just the general fundamental requirement for the highest conversion efficiency of energy.[1, 2] The same as reversible process and quasistatic process, "nondissipative process" is also an ideal process. Any single practically taking place process is a"dissipative process", and the corresponding subfield of thermodynamics is called dissipative thermodynamics. In this way, spiral reactions have been found which provide suitable explanations for the single-way cyclical reactions in "linear regime" and for chemical oscillations in "nonlinear regime". Onsager reciprocal relations and their approximation are discussed. Prigogine dissipative structures and Prigogine's carelessness on thermodynamic coupling are also discussed in this Chapter.

5.1 Dissipative Thermodynamics

As early as the 19th century, some complex systems in which several irreversible processes occur simultaneously, such as the thermoelectric phenomena, were well-known. However, thermodynamic researches were mainly focused on simple systems in the 19th century. Thermodiffusion is the simplest example of complex systems. If a closed container, including a uniform gas mixture of two kinds of molecules A and B, is connected with heat reservoirs of two different temperatures, T_1 and T_2 ($T_1 > T_2$), at both its left end and right end, respectively, then a heat flow passes through the gas mixture. After a certain time, due to the influence of heat flow, two kinds of molecules, A and B, prefer to concentrate at different ends of the container, respectively.[3] That results in a change from the originally uniform gas mixture into a nonuniform gas mixture. Therefore, a stationary state with a certain concentration gradient forms in the container, as shown in Fig. 5.1.

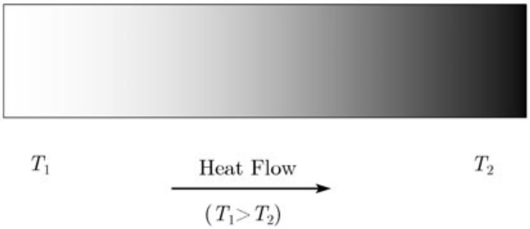


Fig. 5.1 Thermodiffusion

Such a diffusion process from an originally uniform gas mixture into a nonuniform gas mixture is of an opposite direction to an ordinary single diffusion process, so this process is usually called “abnormal diffusion”. The heat flow from high temperature to low temperature is a normal heat conduction. Such a multiprocess system including an obvious influence between different processes is very difficult to treat and understand inside the framework of classical thermodynamics. Thermodiffusion is an objective phenomenon, and it is easy to perform, so it must be a dissipative process with a certain positive entropy production. The difference of temperatures at two ends may be very small, so thermodiffusion may belong to linear dissipative thermodynamics. From the viewpoint of modern thermodynamics this is thermodynamic coupling. The criterion for thermodynamic coupling in the regime of linear dissipative thermodynamics can be written as [$\mathrm{d_i}S_1 < 0$, $\mathrm{d_i}S_2 > 0$ & $\mathrm{d_i}S$ (a little)> 0]. Here, $\mathrm{d_i}S_1$, $\mathrm{d_i}S_2$ and $\mathrm{d_i}S$ are entropy productions of process 1, process 2 and the system, respectively. The criterion can also be written as [$\mathrm{d_i}S_1 < 0$, $\mathrm{d_i}S_2 > 0$ & $\mathrm{d_i}S > 0$] together with a linear relationship between thermodynamic force X and thermodynamic flow J. Here, process 1 is a nonspontaneous “abnormal diffusion” and process 2 is a heat conduction from high temperature to low temperature. Onsager had further dealt with multiprocess systems and deduced his reciprocal relations between coupling coefficients of processes in the multiprocess system.[4] Because of the introduction of linear phenomenological relation between thermodynamic force and thermodynamic flow, and the “assumption of microscopic reversibility” or the assumption of “detailed balancing” during deduction, Onsager reciprocal relations are only approximative relations suitable for the regime of linear dissipative thermodynamics.

In 1958, Belousov reported special kinds of reactions in which organic acid is oxidized by bromic acid under the catalysis of metallic ions. Its particularity is the periodic color change of the whole solution during the reaction progress. For instance, the solution may change alternately between yellow and achromaticity with a minute dimension interval of time, so it is called chemical oscillation. That is an ordered phenomenon with a macroscopic dimension of time. The periodic color change of the solution is due to the periodic change of reaction ion concentrations in the solution, as shown in Fig. 5.2. A little later in 1964, Zhabotinsky reported that ordered patterns of macroscopic spatial dimension can also be produced by these reactions. After then, in the literature all of these reactions are usually called Belousov-Zhabotinsky reactions, or more simply called B-Z reactions. The earliest report on chemical oscillation was Bray’s research in 1921. At this early time their reports underwent skepticism, but later these reactions were well confirmed. The mechanism of these reactions are usually rather complex, and they are hardly, or in no way, to be explained on the basis of classical thermodynamics. Besides B-Z reactions, there are a lot of systems concerned with such kinds of macroscopic spatial and temporal ordered phenomena.

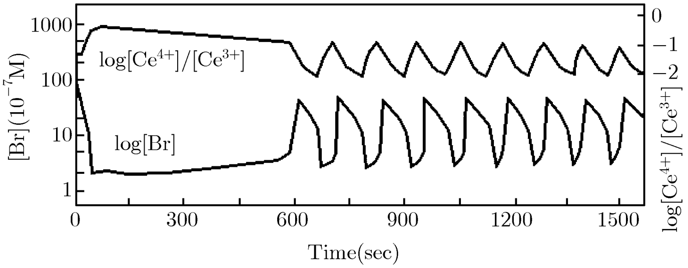


Fig. 5.2 Chemical oscillation. An example of Belousov–Zhabotinsky chemical oscillation reactions. After R.-S. Li (p. 259) [6]

Prigogine made progress in the treatment of these problems. [5, 6] He called these macroscopic spatial and temporal ordered states "dissipative structures", because all of the formation and maintenance of these macroscopic spatial and temporal ordered structures depend on energy dissipation. In particular, there are obvious differences when comparing these macroscopic spatial and temporal ordered structures with microscopic ordered structures usually in molecular or atomic dimension. The latter are usually called "equilibrium structures". Energy dissipation is not necessary for the formation and maintenance of equilibrium structures. For example, liquid water may spontaneously transform into solid state of crystalline ice during temperature decrease to a certain degree. In ice crystals molecules are orderly arranged, but in liquid water molecules are relatively disorderly arranged. Such a kind of ordering change due to interactive force between molecules under low temperature, i.e. the formation of "equilibrium structure", can be reasonably explained in classical thermodynamics. However, the formation of macroscopic spatial and temporal ordered "dissipative structure" under nonlinear conditions may belong to the regime of nonlinear dissipative thermodynamics with a high dissipation of useful work. The key effect is the auto-catalysis effect, or positive feedback effect, which must be involved in this kind of reactions or processes, and has a nonlinear character. The formation of dissipative structures may be as a result of chemical reactions or may be as a result of physical processes, so in thermodynamics the word of "reaction" or "process" is often equally used. From the viewpoint of thermodynamic coupling, the criterion for nonlinear dissipative thermodynamics can be written as $[\mathrm{d_i}S_1 < 0,\ \mathrm{d_i}S_2 > 0\ \&\ \mathrm{d_i}S \gg 0]$. Here, $\mathrm{d_i}S_1$, $\mathrm{d_i}S_2$ and $\mathrm{d_i}S$ are entropy productions of nonspontaneous process(es) 1, spontaneous process(es) 2 and the system, respectively. The criterion can also be written as $[\mathrm{d_i}S_1 < 0,\ \mathrm{d_i}S_2 > 0\ \&\ \mathrm{d_i}S > 0]$ together with a nonlinear relation between thermodynamic force X and thermodynamic flow J.

Figure 5.1 and Fig. 5.2 show two typical examples of experiments. It can be clearly seen in these figures that both thermodiffusion and chemical oscillation may exist and repeat objectively. Both of them must be dissipative processes and belong to the modern dissipative thermodynamics. For the purpose of distinction from classical nonequilibrium thermodynamics, a nomenclature of modern dissipative thermodynamics or dissipative thermodynamics is adopted. It is also relatively opposite to nondissipative thermodynamics, which will be discussed in the later chapters. It is not necessary to change the name of classical nonequilibrium thermodynamics to the name classical dissipative thermodynamics, because the name of classical nonequilibrium thermodynamics is relatively opposite to the name of classical equilibrium thermodynamics.

5.2 Linear Dissipative Thermodynamics and Onsager Reciprocal Relations

According to experiments, the existence of certain linear phenomenological relations in irreversible processes had already been observed many years ago. For instance, Fourier law of heat conduction shows that heat flow in an isotropic medium is directly proportional to the gradient of the temperature (or the gradient of the reciprocal of the temperature) in absolute scale, that is

$$\varTheta = L\nabla\left(\frac{1}{T}\right) = -\frac{L}{T^2}\frac{\partial T}{\partial r} = -\lambda\left(\frac{\partial T}{\partial r}\right). \tag{5.1}$$

Here, $\lambda = \frac{L}{T^2}$ is called the coefficient of heat conduction, and $\varTheta$ is the heat flow. As another example, Fick law of diffusion in binary systems also coincides with this form of linear relationship.

Meanwhile, it should be noted that the discussion of these simple linear relationships is concerned with a system including only one irreversible process. In fact, many experiments show that if there are several irreversible processes in a system, a flow of one irreversible process will depend not only on the corresponding thermodynamic force of the process, but also on the influences of other irreversible processes. That is to say there are some kinds of influences between these irreversible processes. The influence between different irreversible processes is usually called coupling or thermodynamic coupling. For instance, in a binary component gas system, a temperature gradient can lead to not only a heat flow but also a diffusion flow of matter. This is the well-known thermodiffusion phenomenon, as shown in Fig. 5.1. Besides, a concentration gradient in a nonuniform system can lead to not only a flow of matter but also a flow of heat. Therefore, a thermodynamic flow J_k is a function of all other thermodynamic forces X_l in the system:

$$J_k = J_k(\{X_l\}). \tag{5.2}$$

If the situation of several thermodynamic forces $\{X_l\} = 0$ and several thermodynamic flows $J_k(\{X_l\}) = 0$ are taken as the reference point, then $J_k(\{X_l\})$ can be spread as a function at the reference point denoted by subscript zero:

$$\begin{aligned} J_k(\{|X_l\}) = J_k(\{X_{l,\,0}\}) &+ \sum_l \left(\frac{\partial J_k}{\partial X_l}\right)_0 X_l \\ &+ \frac{1}{2}\sum_{l,\,m}\left(\frac{\partial^2 J_k}{\partial X_l X_m}\right)_0 X_l X_m + \cdots. \end{aligned} \tag{5.3}$$

If the system is close to the reference point, in which the thermodynamic forces of all these irreversible processes (X_l, $\cdots$, X_k, $\cdots$) are so small that the high power terms are negligible, only the linear term need be considered, and $J_k(\{X_{l,\,0}\}) = 0$; then a linear phenomenological relation can be obtained as

$$J_k = \sum_l L_{kl} X_l, \tag{5.4}$$

where, $L_{kl} = \left(\frac{\partial J_k}{\partial X_l}\right)_0$. The phenomenological coefficients L_{kk} represent the relation between the thermodynamic forces X_k and the corresponding thermodynamic flow J_k, while the phenomenological coefficients L_{kl} $(k \neq l)$ represent the coupling relationship between different irreversible processes. The phenomenological coefficients L_{kl} could be concerned with pressure, temperature, concentrations etc. but not related to the change rate of these parameters.

As mentioned above, the linear phenomenological relation is observed in research labs, but is not deduced from the basic laws of thermodynamics. Even by a backward step, it might be introduced as an assumption. If once the phenomenological laws are accepted, some properties of the phenomenological coefficients L_{kl} can be obtained by thermodynamic methods from other natural principles.

Based on the introduction of the linear phenomenological relation as approximation or assumption, if a system including two simultaneous irreversible processes is under consideration, then the interaction (i.e. coupling effect or cross effect) can be written as

$$J_1 = L_{11} X_1 + L_{12} X_2, \tag{5.5}$$

$$J_2 = L_{21} X_1 + L_{22} X_2. \tag{5.6}$$

The entropy production per unit time of the system will be

$$\begin{aligned} \sigma &= \sum_k J_k X_k = J_1 X_1 + J_2 X_2 \\ &= L_{11} X_1^2 + (L_{12} + L_{21}) X_1 X_2 + L_{22} X_2^2. \end{aligned} \tag{5.7}$$

According to the second law of thermodynamics, $\sigma \geqslant 0$ must be satisfied, and that is independent of the selected values of X_1 and X_2. Apart from $X_1 = X_2 = 0$ and $\sigma = 0$, the selected values of X_1 and X_2 must satisfy the following sufficient and necessary conditions required in algebra:

$$L_{11} > 0,\ L_{22} > 0 \text{ and } (L_{12} + L_{21})^2 < 4 L_{11} L_{22}. \tag{5.8}$$

That is to say, these phenomenological coefficients are not arbitrary, but subordinate to certain limitations and principles.

Besides the second law of thermodynamics, these phenomenological coefficients are also limited by the Curie's space symmetry principle. The Curie's space symmetry principle shows that the symmetric elements of a macroscopic cause are always less than that of the produced effect in an isotropic medium. Prigogine introduced it into the treatment of irreversible processes, and concluded that the symmetric elements of a thermodynamic force must not be more than that of the thermodynamic flow which is coupled with it. That is called Currie-Prigogine's principle.[6] For instance, a chemical reaction will not be able to drive a flow of matter or a flow of heat. It should be noted that the premise of the conclusion is for reactions in an isotropic system. For reactions taking place on a film or on a surface the limitation will not be valid. Several active diffusions (i.e. diffusions contrary to the concentration gradient) taking place in biological membranes are examples without such limitations.

In addition, if the phenomenological coefficients also subordinate the limitation of the "assumption of microscopic reversibility" or the assumption of "detailed balancing", then the Onsager's reciprocal relations could be deduced. That is, there is a symmetrical character between the linear phenomenological coefficients. [1], [4]-[8] Its mathematical expression is

$$L_{kk'} = L_{k'k}. \tag{5.9}$$

Its written statement is that if the flow of the kth irreversible process J_k is influenced by the force of the k' th irreversible process $X_{k'}$, then the flow of the k'th irreversible process $J_{k'}$ must also simultaneously be influenced by the force of the kth irreversible process X_k. That is, the two coupling coefficients characterizing the influences on each other between these two irreversible processes must be equal.

In some textbooks, it has been told that Onsager reciprocal relations Eq. (5.9) bring us convenience in research on multiprocess systems. For a system including n simultaneous irreversible processes, it should originally be necessary to solve n^2 linear coefficients, but now only $\frac{n(n+1)}{2}$ independent phenomenological coefficients should be solved. If n is a very big number, a half of the work will be reduced. In this way it is possible to get some cross phenomenological coefficients, which are difficult to be directly observed and measured through some cross phenomenological coefficients, which are easy to be observed and measured. However, in fact, at the time of Onsager, and even up to now, research objects are mainly concerned with system including two simultaneous irreversible processes. For a system including two simultaneous irreversible processes the number of phenomenological coefficients, which should be observed and measured, will be reduced from the original 4 to 3. Among them, the coupling phenomenological coefficients will be reduced from the original 2 to 1. At the same time note that Onsager reciprocal relations are not rigorous thermodynamic relations.

5.3 Cyclical Reactions

A cyclical reaction will be discussed as an example as follows. [1], (p. 88) [6], (p. 609) [7] If there is a reaction cycle in a reaction system, as shown in Fig. 5.3, then the formulations of the three reaction rates and affinities are

$$r_1 = k_1(\mathrm{A}) - k_{-1}(\mathrm{B}),\ A_1 = \mu_\mathrm{A} - \mu_\mathrm{B}, \tag{5.10}$$

$$r_2 = k_2(\mathrm{B}) - k_{-2}(\mathrm{C}),\ A_2 = \mu_\mathrm{B} - \mu_\mathrm{C}, \tag{5.11}$$

and

$$r_3 = k_3(\mathrm{C}) - k_{-3}(\mathrm{A}),\ A_3 = \mu_\mathrm{C} - \mu_\mathrm{A}. \tag{5.12}$$

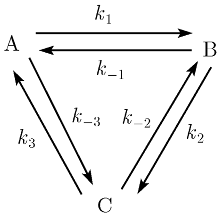


Fig. 5.3 Cyclical Reactions

(A), (B) and (C) represent the concentrations of substances A, B and C, respectively. Obviously, only two of the formulations are independent among the three expressions for the affinities listed above. That is

$$A_3 = -(A_1 + A_2). \tag{5.13}$$

Therefore, the entropy production per unit time of the whole system can be written as

$$\sigma = \frac{1}{T}\sum_{k=1}^{3} A_k r_k = \frac{1}{T}\left[A_1 (r_1 - r_3) + A_2 (r_2 - r_3)\right]. \tag{5.14}$$

If A_1/T and A_2/T in the expression are taken as two independent thermodynamic forces, X_1 and X_2, respectively, then two corresponding thermodynamic flows are J_1 and J_2, respectively:

$$X_1 = A_1/T,\ J_1 = r_1 - r_3, \tag{5.15}$$

$$X_2 = A_2/T,\ J_2 = r_2 - r_3. \tag{5.16}$$

If the system is in equilibrium,

$$\mu_{\mathrm{A},\,0} = \mu_{\mathrm{B},\,0} = \mu_{\mathrm{C},\,0} = 0, \tag{5.17}$$

then

$$A_{1,\,0} = A_{2,\,0} = A_{3,\,0} = 0, \tag{5.18}$$

$$r_{1,\,0} = r_{2,\,0} = r_{3,\,0} = 0. \tag{5.19}$$

During the deduction, an “additional restriction” had been introduced by Onsager. That is,

> ···when the equilibrium is reached each individual reaction must balance itself. They require that the transition A → B must take place just as frequently as the reverse transition B → A etc.

That should be the “assumption of microscopic reversibility” or the assumption of “detailed balancing” as called by Onsager. In other words, any forward and backward mean rates in a microscopic process during equilibrium must equal each other.

$$k_1(\mathrm{A})_0 = k_{-1}(\mathrm{B})_0,\ k_2(\mathrm{B})_0 = k_{-2}(\mathrm{C})_0,\ k_3(\mathrm{C})_0 = k_{-3}(\mathrm{A})_0, \tag{5.20}$$

where, $(\mathrm{A})_0$, $(\mathrm{B})_0$ and $(\mathrm{C})_0$ represent the concentrations of the three components in equilibrium, respectively.

If now the system deviates a little from its equilibrium state, suppose that the concentrations of the three components are

$$(\mathrm{A}) = (\mathrm{A})_0 + x,\ (\mathrm{B}) = (\mathrm{B})_0 + y,\ (\mathrm{C}) = (\mathrm{C})_0 + z, \tag{5.21}$$

respectively, and the following requirements are satisfied, i.e.

$$\frac{|x|}{(\mathrm{A})_0} \ll 1,\ \frac{|y|}{(\mathrm{B})_0} \ll 1,\ \frac{|z|}{(\mathrm{C})_0} \ll 1, \tag{5.22}$$

then

$$r_1 = k_1 x - k_{-1} y,\ r_2 = k_2 y - k_{-2} z,\ r_3 = k_3 z - k_{-3} x, \tag{5.23}$$

and

$$A_1 = \mu_A - \mu_B = A_{1,0} + RT\ln\left[1+\frac{x}{(\mathrm{A})_0}\right] - RT\ln\left[1+\frac{y}{(\mathrm{B})_0}\right] \tag{5.24}$$
$$\approx RT\left[\frac{x}{(\mathrm{A})_0} - \frac{y}{(\mathrm{B})_0}\right] = \frac{RT}{k_1(\mathrm{A})_0}(k_1x - k_{-1}y).$$

Therefore,

$$r_1 = \frac{k_1(\mathrm{A})_0}{RT}A_1. \tag{5.25}$$

By a similar derivation the following equations can be obtained.

$$r_2 = \frac{k_2(\mathrm{B})_0}{RT}A_2 \tag{5.26}$$

and

$$r_3 = \frac{k_3(\mathrm{C})_0}{RT}A_3 = -\frac{k_3(\mathrm{C})_0}{RT}(A_1 + A_2). \tag{5.27}$$

Therefore, the phenomenological relations are

$$J_1 = r_1 - r_3 = \frac{k_1(\mathrm{A})_0 + k_3(\mathrm{C})_0}{R}\cdot\frac{A_1}{T} + \frac{k_3(\mathrm{C})_0}{R}\cdot\frac{A_2}{T} \tag{5.28}$$
$$= L_{11}X_1 + L_{12}X_2$$

and

$$J_2 = r_2 - r_3 = \frac{k_3(\mathrm{C})_0}{R}\cdot\frac{A_1}{T} + \frac{k_2(\mathrm{B})_0 + k_3(\mathrm{C})_0}{R}\cdot\frac{A_2}{T} \tag{5.29}$$
$$= L_{21}X_1 + L_{22}X_2.$$

Where the phenomenological coefficients are

$$L_{11} = \frac{k_1(\mathrm{A})_0 + k_3(\mathrm{C})_0}{R}, \tag{5.30}$$

$$L_{12} = \frac{k_3(\mathrm{C})_0}{R} = L_{21}, \tag{5.31}$$

$$L_{22} = \frac{k_2(\mathrm{B})_0 + k_3(\mathrm{C})_0}{R}. \tag{5.32}$$

In this way, it has been confirmed that Onsager reciprocal relations are valid in the system. During the deduction of the relations, the assumption of “detailed balancing”, i.e. Eq. (5.20), and the conditions of a slight deviation of the system from equilibrium, i.e. Eq. (5.22), have been adopted. Therefore, that is also a suitable range for Onsager reciprocal relations. The relations in general form ($L_{kk'} = L_{k'k}$) were deduced by Onsager in 1931, and were regarded as fundamental for linear dissipative thermodynamics.

However, if the discussed cyclical reaction system, as shown in Fig. 5.3, were rewritten in Fig. 5.4, that would be a single-way cyclical reaction system. If such a system were in equilibrium, there would be nothing violating the thermodynamic laws, but such a reaction does violate the "assumption of microscopic reversibility" or the assumption of "detailed balancing". Therefore, it does not practically exist. Onsager had carefully pointed out that the "assumption of microscopic reversibility" or the assumption of "detailed balancing" was out of thermodynamics, and might be a principle independent of thermodynamics. The conclusion that the single-way cyclical equilibrium does not exist is not coming from thermodynamics. Therefore, the single-way cyclical equilibrium may be a form of thermodynamic extremum. This is just as "nondissipative process" does not practically exist, but in thermodynamics it is very important. Discussion on it will be continued in the following Section 5.5 .

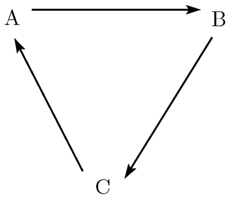


Fig. 5.4 A single-way cyclical reaction system

5.4 Entropy Production Minimization Principle

Besides Onsager reciprocal relations another principle of linear dissipative thermodynamics may be the entropy production minimization principle proposed by Prigogine in 1945. That is, in a linear nonequilibrium system the stationary state with respect to external constrained conditions may be a state with a minimum of entropy production. (p. 96) [6, 9], (p. 76) [10]

For any system without external constraint (such as any isolated system), free development will result in an equilibrium state. If the state function of the entropy is taken as a criterion, the entropy of the system will reach its maximum value. For an isothermal and isobaric change of a closed system it is more convenient to use the Gibbs free energy as a criterion. The system has a minimum of Gibbs free energy during equilibrium. The Gibbs free energy minimization principle is the most useful under general constant temperature and pressure conditions. In equilibrium states, no macroscopic process and no irreversible process takes place inside the system, so the entropy production of equilibrium system equals zero. Be careful, during equilibrium entropy production of the system is not maximum.

Under an externally constrained condition, the system cannot reach an equilibrium state but a stationary state independent of time instead. For instance, if both sides of a closed binary gaseous system are connected with heat reservoirs at different temperatures, respectively, there will be a definite temperature gradient in the system. At the same time a concentration gradient will be set up in the system. This is the well-known thermodiffusion phenomenon. In the final state of the system there is a steady heat flow and a fixed concentration gradient, but the mass flow of diffusion will be zero. In the stationary state of system the entropy production will have a minimum value. For example, a simple demonstration may be deduced as follows.

According to the general expression the entropy production per unit time may be written as

$$\sigma = J_1 X_1 + J_2 X_2. \tag{5.33}$$

The subscripts 1 and 2 represent heat conduction and diffusion processes, respectively. It is assumed that the heat conduction and the diffusion are satisfied by the linear relation, and that Onsager reciprocal relations are valid, so that

$$J_1 = L_{11}X_1 + L_{12}X_2 \tag{5.34}$$

and

$$J_2 = L_{21}X_1 + L_{22}X_2. \tag{5.35}$$

$L_{12} = L_{21}$,

$$\sigma = L_{11}X_1^2 + 2L_{12}X_1X_2 + L_{22}X_2^2. \tag{5.36}$$

During the free development of the diffusion force under the constrained condition of constant heat conduction force X_1, the change of the entropy production can be obtained by the partial differential of σ with respect to X_2:

$$\frac{\partial \sigma}{\partial X_2} = 2(L_{21}X_1 + L_{22}X_2) = 2J_2. \tag{5.37}$$

When a stationary state is reached, the diffusion flow equals zero, i.e. $J_2 = 0$, so

$$\frac{\partial \sigma}{\partial X_2} = 0. \tag{5.38}$$

That is, for a stationary state the entropy production per unit time should be an extreme value with respect to the change of the diffusion force, and $\frac{\partial^2 \sigma}{\partial X_2^2} > 0$, so the extreme value must be a minimum. The general demonstration of this principle on the basis of the variational method will not be further discussed here. It should be noted that during the deduction of the entropy production minimization principle, not only Onsager reciprocal relations but also the assumption of constant phenomenological coefficients had been used, so that the domain of validity of the Prigogine entropy production minimization principle is stationary states of the linear dissipative system.

Entropy production minimization principle shows that the general tendency of development of systems in the linear regime of nonequilibrium states is the decrease of entropy production. If the equilibrium state has been prevented from being reached due to the externally constrained condition, the system has to choose an entropy production minimization state, i.e. a stationary state. An equilibrium state at constant temperature and pressure corresponds to a Gibbs free energy minimization state, and there is no dissipation of Gibbs free energy. However, in general, for a system in its stationary state there may be of a dissipation of the Gibbs free energy, but it may be a dissipation minimization state under a definite externally constrained condition. Therefore, it can be found that the change from an equilibrium state into a stationary state is continuous. Equilibrium state may also regard as a special stationary nonequilibrium state with zero entropy production or with zero dissipation of Gibbs free energy. It should be noted that if an externally constrained condition exists, the state with zero entropy production or with zero dissipation of Gibbs free energy may also be an ideal nonequilibrium nondissipative state.

5.5 Approximation of Onsager Reciprocal Relations

It was known very early that if two or more irreversible transport processes (heat conduction, electrical current and diffusion) take place simultaneously in a thermodynamic system, then these processes may interfere with each other. Thus an electric current in a circuit that consists of different metallic conductors will in general cause evolution or absorption of heat at the junctions (Peltier effect). Conversely, if the junctions are maintained at different temperatures an electromotive force will usually appear in the circuit, the thermoelectric force; the flow of heat has a tendency to carry the electricity.

In such cases one may naturally suspect reciprocal relations by analogy to reciprocal relations which connect forces and displacements in the equilibrium theory of mechanics and in thermodynamics. Reciprocal relations had been proposed and discussed by many writers. The earliest of them all was W. Thomson (Lord Kelvin) who dealt with thermoelectric phenomena.[11] Thomson reciprocal relations were written in the form of resistances, so the relations for interference between these two processes were

$$X_1 = R_{11}J_1 + R_{12}J_2, \tag{5.39}$$

$$X_2 = R_{21}J_1 + R_{22}J_2. \tag{5.40}$$

Here, J_1 and J_2 represent the electric current and the heat flow, respectively. The current is driven by the electromotive force, X_1, while in corresponding units the heat flow is driven by the “force”: $X_2 = -\frac{1}{T}\mathrm{grad}T$, where T denotes the absolute temperature, R_1 is the electrical resistance and R_2 is a “heat resistance”. Therefore, Thomson reciprocal relations were:

$$R_{12} = R_{21}. \tag{5.41}$$

Thomson arrived at this relation from thermodynamic reasoning, but he had to make one additional assumption, namely:

> The electromotive forces produced by inequalities of temperature in a circuit of different metal, and the thermal effect of electric current circulating in it, are subject to the laws which would follow from the general principles of the thermodynamic theory of heat if there were no conduction of heat from one part of the circuit to another.

Thomson thought this assumption very plausible. Even so, he cautiously considered his reciprocal relation Eq. (5.41) a conjecture to be confirmed or refuted by experiment since it could not be derived entirely from fundamental principles known at that time. However, as written by Onsager himself in 1931:[4]

> At present Thomson’s relation is generally accepted, because it has been confirmed within the limits of error of the best measurements.

As early as 1876, Helmholtz[12] first introduced the form of conductance, i.e. Eq. (5.5) $J_1 = L_{11}X_1 + L_{12}X_2$ and Eq. (5.6) $J_2 = L_{21}X_1 + L_{22}X_2$. He had also got $L_{12} = L_{21}$. That was the expression of Helmholtz reciprocal relation. Onsager reciprocal relations, Eq. (5.9) $L_{ij} = L_{ji}$, were just the extended form of Helmholtz reciprocal relation. Just as Onsager said:[4, 13]

> Helmholtz has derived the reciprocal relation. $L_{12} = L_{21}$. the experiments confirm the result, Nernst has given a kinetic derivation from assumptions that are somewhat specialized, and the theorem is generally accepted.

The deduction of Onsager reciprocal relations was based on more general rules of irreversible processes, and the "assumption of microscopic reversibility" or the assumption of "detailed balancing" in chemical kinetics. The resultant relations, $L_{kk'} = L_{k'k}$ are of more general form in comparison with Thomson reciprocal relations and Helmholtz reciprocal relation. That was Onsager's contribution to the development of thermodynamics for complex systems (or modern thermodynamics).

Obviously, Onsager reciprocal relations are different from Thomson reciprocal relations, but are completely the same as Helmholtz reciprocal relation, if only two irreversible processes were under consideration. In fact, all of these reciprocal relations are approximate. Many experimental data agree with Onsager reciprocal relations, i.e. $L_{qe}/L_{eq} = 1$. However, deviation of $L_{qe}/L_{eq} = 0.77$ was also found, as shown in Table 5.1.[5] Thomson reciprocal relations, Helmholtz reciprocal relation and Onsager reciprocal relations were not all deduced strictly from thermodynamics. The physical idea of them was conjecture that the interaction coefficient (or coupling coefficient) of the *k*th irreversible process on the *k*′th irreversible process would equal or approach the interaction coefficient (or coupling coefficient) of the *k*′th irreversible process on the *k*th irreversible process.

Table 5.1 Comparison between Onsager reciprocal relations ($L_{qe}/L_{eq} = 1$) and experiments

Thermocouple	T (°C)	L_{qe}/L_{eq}
Cu – Al	15.8	0.77
Cu – Ni	0	0.930
Cu – Ni	14	0.976
Cu – Fe	0	1.000
Cu – Bi	20	1.08
Fe – Ni	16	1.06
Fe – Hg	18.4	1.004

In comparison with reciprocal relations, the influences or interaction between irreversible processes in this book are directly and clearly expressed as thermodynamic coupling, i.e. the mathematical expression of $[\mathrm{d_i}S_1 < 0,\ \mathrm{d_i}S_2 > 0\ \&\ \mathrm{d_i}S \geqslant 0]$. Here, entropy production of system is the sum of entropy production of internal processes, i.e. $\mathrm{d_i}S = (\mathrm{d_i}S_1 + \mathrm{d_i}S_2)$. That is a quantitative relation. For isothermal isobaric systems, the mathematical expression of thermodynamic coupling is $[(\mathrm{d}G_1)_{T,p} > 0,\ (\mathrm{d}G_2)_{T,p} < 0\ \&\ (\mathrm{d}G)_{T,p} \leqslant 0]$. Here, Gibbs free energy change of system is the sum of Gibbs free energy change of internal processes, i.e. $(\mathrm{d}G)_{T,p} = [(\mathrm{d}G_1)_{T,p} + (\mathrm{d}G_2)_{T,p}]$. Therefore, the relation for thermodynamic coupling is a very clear and quantitative relation of energy. That is the difference between thermodynamic coupling in this book and "coupling" in Onsager reciprocal relations.

Besides, there was another confusion for a long time in dissipative thermodynamics (or called nonequilibrium thermodynamics in the past), i.e. a confusion between linear region and "close-to-equilibrium (or nearly in equilibrium)". For instance, Onsager said in his famous paper[4]

> In order to obtain proportionality between the "forces" X and the displacements x we must limit ourselves to the consideration of cases where the system is nearly in equilibrium, $\cdots$.

The former was linear relation, but the later became “nearly in equilibrium”.

Meanwhile, Onsager often discussed nonequilibrium states around thermodynamic equilibrium. In fact, under nonequilibrium conditions, systems may not be able to reach equilibrium. For instance, in Fig. 5.1 Thermodiffusion, if temperature gradient ($T_1 > T_2$) exists, neither A molecules nor B molecules will be able to reach uniform equilibrium state. In other words, no extremum exists for both state functions of A and B, so no “close-to-equilibrium” could be concerned. Therefore, the state “close-to-equilibrium” must be around extremum of state function, but “nonequilibrium” may be no relation with extremum. “Linear nonequilibrium” and “close-to-equilibrium” are two different concepts in thermodynamics. “Nonlinear nonequilibrium” and “far-from-equilibrium” are two different concepts in thermodynamics as well. We should not confuse one with another between them.

Besides, note that the starting point of these reciprocal relations was of some problems even at their time. First, they believed that any irreversible process would belong to inequality part of the second law of thermodynamics, and that it was impossible to get an equality directly from the second law of thermodynamics (now we know this viewpoint is incorrect, and this will be discussed in the next chapter about a new field of nondissipative thermodynamics). Second, from analogy to mechanics they believed that interactions between irreversible processes were equal (at least approximate). Third, all of them introduced empirical (out-of thermodynamics) phenomenological relations for interaction or coupling effect, and then approached approximation. Fourth, all of these different reciprocal relations approximately agreed with or were close to experimental data, but all of them are “quasi-thermodynamic theories”, because out-of-thermodynamics assumptions had been introduced in their deductions. Among them Onsager reciprocal relations were more general, so more widely accepted.

To sum up, both Thomson reciprocal relations and Onsager reciprocal relations (or Helmholtz reciprocal relations) are approximately supported by experimental data, so Onsager reciprocal relations should never be regarded, with absolute meaning, as the fundamental laws of thermodynamics, so-called the fourth law of thermodynamics.[1] Meanwhile, Prigogine deduced entropy production minimization principle on the basis of Onsager reciprocal relations, so the entropy production minimization principle should not be regarded as a fundamental law of thermodynamics with absolute meaning either. The general theorems are the extended Carnot theorem and dissipation (or entropy production) decrease theorem. These theorems are suitable for both nondissipative thermodynamics and dissipative thermodynamicsthe (including linear and nonlinear dissipative thermodynamics). We may even say, these theorems are suitable for the whole thermodynamics discipline, as discussed in the first chapter and in the last chapter.

5.6 Nonlinear Dissipative Thermodynamics and Prigogine Dissipative Structures

Traditional nonlinear nonequilibrium thermodynamics, or traditional thermodynamics of

---

[1] http://encyclopedia.farlex.com/Onsager,+Lars. Cited 28 Apr 2008

nonlinear irreversible processes, should actually be called nonlinear dissipative thermodynamics in a modern classification of thermodynamics. In this regime, systems may be quite different from equilibrium systems, and there are no general extremum principles that predict the state to which the system will evolve. In contrast to simple uncoupling systems, which evolve to a state that minimizes Gibbs free energy, nonlinear dissipative systems can evolve unpredictably, and the states may not always be unique. For a given set of nonequilibrium conditions, it is often possible to have more than one state. Which one of these states a particular system will evolve to is, in general, not predictable. The new states thus attained may often be "ordered states" that possess spatiotemporal organization. Patterns in fluid flow, inhomogeneities in concentrations exhibiting geometrical patterns with great symmetry, or periodic variations of concentrations are examples of such ordered states. This phenomenon is usually called "self-organization" and since it is a result of fluctuation is referred to as "order through fluctuations".

In nonlinear dissipative systems, oscillating concentrations and geometrical concentration patterns can be a result of chemical reaction and diffusion. That was called Turing structures (refer to Sect. 5.10 of this chapter.[5]) The same dissipative processes in a closed or isolated system wipe out inhomogeneities and drive the system to a stationary, timeless homogeneous state of equilibrium. Since the creation and maintenance of organized nonequilibrium states are due to dissipative processes, Prigogine called them "dissipative structures". (p. 154)[6], (p. 124) [10] Roughly speaking, the formation of a macroscopic spatiotemporal ordering process is a negative entropy production process, which takes place under the drive of a strong enough positive entropy production dissipative process. That is also a masterpiece of thermodynamic coupling in the nonlinear dissipative region. In kinetic mechanism, the macroscopic spatiotemporal ordering dissipative structures may be formed through a positive feedback or auto-catalysis to escalate local microscopic fluctuation.

5.7 Bérnard Pattern

In 1900 the famous experiment, "Bérnard Pattern", was performed by Bérnard and reported in his doctoral thesis. (p. 39, 286)[6], (p. 10–12)[14] In this experiment, an adverse temperature gradient is maintained in a thin horizontal layer of fluid by heating the underside. Due to thermal expansion, the fluid at the bottom will be lighter than the fluid at the top. This top-heavy arrangement is potentially unstable. There will be a tendency on the part of the fluid to redistribute itself and overcome the instability in its arrangement. However, this natural tendency is inhibited by its own viscosity. When the adverse temperature gradient exceeds a certain threshold value, the fluid starts to organize itself to show some kind of macroscopic regular patterns depending on various boundary conditions. For instance, (a) if the upper surface of fluid is free in the air, then the corresponding Bérnard patterns are a regular hexagonal cellular pattern, as shown in Fig. 5.5 (a); (b) if the upper surface of fluid is covered with a solid plate in a round container, then the corresponding Bérnard patterns are a concentric circle pattern, as shown in Fig. 5.5 (b); or (c) if the upper surface of fluid is covered with a solid plate but in a square container, then the corresponding Bérnard patterns are a parallel line pattern, as shown in Fig. 5.5 (c), respectively.[15]

The phenomenon can also be called “self-organization”. Figure 5.6 shows a cross section of the fluid at the first moment of such a “self-organization”. The exact time and the exact place of self-organization starting may be ambiguous due to fluctuation, and may also be induced manually by a turbulence at a special time and place. However, once the first cell — “Bérnard’s roll” forms, the dissection of the layer into cells follows. This is the key

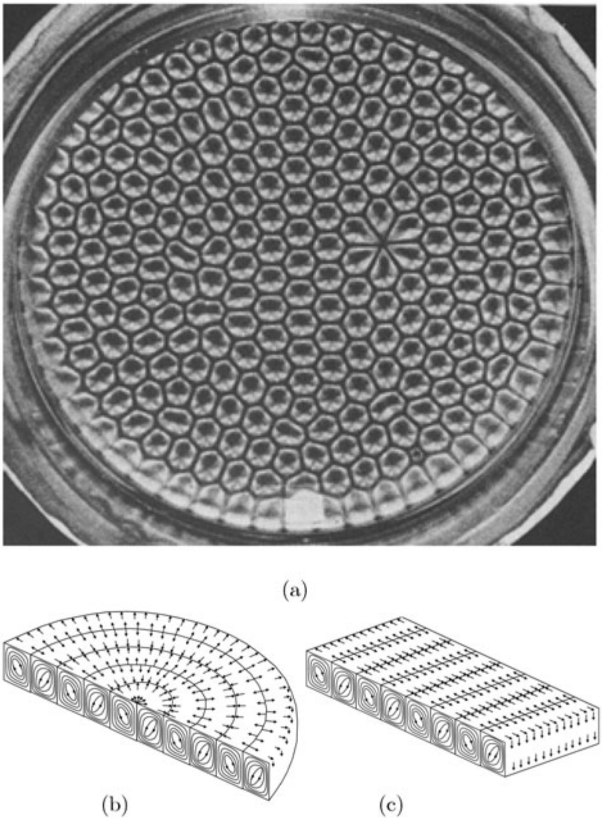


Fig. 5.5 (a) A regular hexagonal cellular pattern of Bérnard’s pattern, (b) A concentric circle pattern of Bérnard’s pattern, (c) A parallel line pattern of Bérnard’s pattern. (a) is reprinted from http://www.meta-synthesis.com/webbook/24_complexity/complexity3.html. Cited 1 Dec 2008 and originally from E.L. Koschmieder, *Adv. Chem. Phys.* 26, 177 (1974)

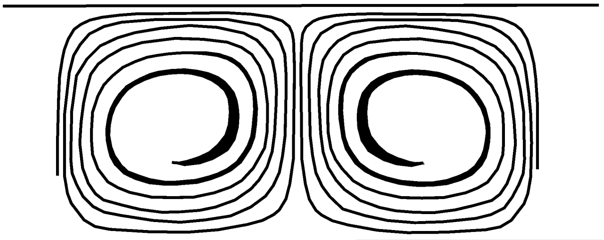


Fig. 5.6 Cross-section of Self-organized “Bérnard’s roll” at the first moment of its formation (p. 40)[16]

factor in forming a dissipative structure, and is called the positive feedback effect (or autocatalysis effect in chemical reactions). There are also motions inside the cells: of ascension at the center, and of descent at the boundaries with the adjoining cells. The detailed hydrodynamic equations can be written out, but they involve mainly hydrodynamics or kinetics and will not be discussed in detail here.

5.8 Laser Emission

The laser (light amplification by stimulated emission of radiation) appeared in the 1960s. It is also a typical example of a macroscopically ordered dissipative structure. (p. 63)[16] The amplification effect (i.e. positive feedback effect) of stimulated emission is the key factor in producing laser. Here, the light intensity I will be taken as a parameter for discussion. Under normal conditions, atoms emit and absorb photons at random, so $I \approx 0$. During stimulated emission $I \neq 0$, some factors related to $\frac{\mathrm{d}I}{\mathrm{d}t}$ should be considered.

1. The light intensity of stimulated emission is directly proportional to the light intensity of incidence. That is, the increase of light intensity due to stimulation emission will be directly proportional to
$$DI, \tag{5.42}$$
D is called a gain coefficient. Obviously, D is related with the atom population number at the excited level N_i. Increasing N_i will cause D to increase as well.
2. Atom stimulation needs absorption of photons, and results in a decrease of the light intensity I and a negative value of $\frac{\mathrm{d}I}{\mathrm{d}t}$. That is, the increase of light intensity due to atom stimulation will be also directly proportional to
$$-A_c I. \tag{5.43}$$
A_c is called a wear coefficient (and includes other effects that decrease the light intensity). Combining these two factors, the equation for the change in light intensity $\frac{\mathrm{d}I}{\mathrm{d}t}$ can be written as
$$\frac{\mathrm{d}I}{\mathrm{d}t} = DI - A_c I, \tag{5.44}$$
This is a linear equation. If $D > A_c$, the light intensity would increase infinitely. This is physically impossible to realize. It is prevented by the limitation of the atom population number at the excited level. It can be called the saturation effect, and is nonlinear. Suppose that the nonlinear effect can be represented by the simple nonlinear term
$$-\beta I^2, \tag{5.45}$$
β is the called saturation coefficient. Now Eq. (5.44) should be rewritten as
$$\frac{\mathrm{d}I}{\mathrm{d}t} = (D - A_c)I - \beta I^2, \tag{5.46}$$
This is the frequently used light intensity equation for the laser.

For Eq. (5.46) there are two stationary solutions:

$$I_1 = 0 \text{ and } I_2 = (D - Ac)/\beta, \tag{5.47}$$

If the whole system is in equilibrium or "near-equilibrium", the atom population number at the excited level will be much less than that at the ground level, so $D < A_c$. In such a case, I_2 would be less then zero. This solution for I_2 is meaningless because light intensity is never negative. Therefore, Eq. (5.46) has only one solution with no laser. Based on linear stability analysis, this solution of disorder is stable.

If the intensity of the photon pump increases, D will increase with the atom population number at the excited level. When the intensity of the photon pump is strong enough for a reversed atom population, i.e. the gain coefficient is greater than the wear coefficient $(D > A_c)$, the nonzero stationary solution I_2 becomes positive and meaningful.

Based on linear stability analysis, the solution $I_1 = 0$ becomes unstable, while the solution $I_2 = (D - Ac)/\beta$ becomes a stable solution. The result is illustrated in Fig. 5.7. When $D < A_c$, there is a solution of $I_1 = 0$. When $D > A_c$, there are two solutions: a stable solution (solid line) of $I_2 = (D - A_c)/\beta$ and an unstable solution (dashed line) of $I_1 = 0$.

The laser example also shows that all these photons are emitted with the same frequency, the same direction and the same phase. This is a further example of self-organization.

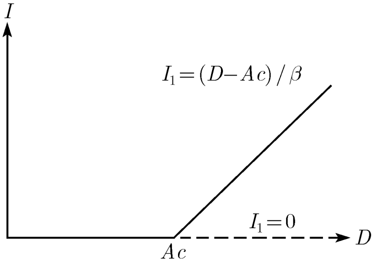


Fig. 5.7 Relation between the light intensity I and the gain coefficient D. The solid line represents the stable solution, while the dashed line represents the unstable solution. After Xiao-Feng Shen (p.65)[16]

5.9 Chemical Oscillation and "Brusselator"

Chemical oscillation shows another kind of dissipative structure, which illustrates how the breaking of time-translation symmetry leads to oscillatory behavior. It is difficult to find an explanation in classical thermodynamics. In 1968, Prigogine and his coworkers proposed a typical kinetic model, called Brusselator or trimolecular model. On the basis of this model, the evolution mechanism of a nonequilibrium system to be unstable and to form an oscillation can be well explained. [2]

First of all, the following simplified reaction set is under consideration:(p. 439)[5]

$$\mathrm{A} \xrightarrow{k_1} \mathrm{X}, \tag{5.48}$$

---

[2] This Section 5.9 is cited mainly from D. Kondepudi, I. Prigogine, *Modern Thermodynamics — From Heat Engines to Dissipative Structures* (John Wiley & Sons, New York, 1998) pp. 439–440

$$\mathrm{B}+\mathrm{X}\xrightarrow{k_2}\mathrm{Y}+\mathrm{D}, \tag{5.49}$$

$$2\,\mathrm{X}+\mathrm{Y}\xrightarrow{k_3}3\,\mathrm{X}, \tag{5.50}$$

$$\mathrm{X}\xrightarrow{k_4}\mathrm{E}, \tag{5.51}$$

The whole reaction is A + B → D + E. It is assumed that the concentrations of reactants A and B are maintained at a desired nonequilibrium value through appropriate flows. The products of D and E are removed as they are formed. Note that, once the intermediate X is produced in the first step of reaction, it will be produced more quickly by the second and the third steps of reaction. That is the characteristics of nonlinear auto-catalyst reaction. It is also assumed that the reaction occurs in a solution that is well stirred, hence homogeneous, and all the reverse reactions are so slow they can be neglected. Then, the following nonlinear kinetic rate equations for species X and Y can be obtained:

$$\frac{\mathrm{d[X]}}{\mathrm{d}t}=k_1[\mathrm{A}]-k_2[\mathrm{B}][\mathrm{X}]+k_3[\mathrm{X}]^2[\mathrm{Y}]-k_4[\mathrm{X}]\equiv \mathrm{Z}_1, \tag{5.52}$$

$$\frac{\mathrm{d[Y]}}{\mathrm{d}t}=k_2[\mathrm{B}][\mathrm{X}]-k_3[\mathrm{X}]^2[\mathrm{Y}]\equiv \mathrm{Z}_2. \tag{5.53}$$

It can easily be verified that the stationary solutions to these nonlinear equations are

$$[\mathrm{X}]_\mathrm{S}=\frac{k_1}{k_4}[\mathrm{A}] \text{ and } [\mathrm{Y}]_\mathrm{S}=\frac{k_2k_4}{k_3k_1}\frac{[\mathrm{B}]}{[\mathrm{A}]}. \tag{5.54}$$

The stability of the stationary state depends on the eigenvalues of the Jacobian matrix

$$\begin{bmatrix} \dfrac{\partial \mathrm{Z}_1}{\partial \mathrm{X}} & \dfrac{\partial \mathrm{Z}_1}{\partial \mathrm{Y}} \\ \dfrac{\partial \mathrm{Z}_2}{\partial \mathrm{X}} & \dfrac{\partial \mathrm{Z}_2}{\partial \mathrm{Y}} \end{bmatrix}. \tag{5.55}$$

The explicit form of the Jacobian matrix is

$$\begin{bmatrix} k_2[\mathrm{B}]-k_4 & k_3[\mathrm{X}]_\mathrm{S}^2 \\ -k_2[\mathrm{B}] & -k_3[\mathrm{X}]_\mathrm{S}^2 \end{bmatrix}=\Lambda. \tag{5.56}$$

The stationary state Eq. (5.54) becomes unstable when the real part of the eigenvalues of Eq. (5.55) become positive. The eigenvalue equation or characteristic equation of a matrix Λ, whose solutions are the eigenvalues, is

$$\mathrm{Det}[\Lambda-\lambda I]=0. \tag{5.57}$$

"Det" represents the determinant. For a 2 × 2 matrix such as Eq. (5.56), the characteristic equation is

$$\lambda^2-(\Lambda_{11}+\Lambda_{22})\lambda+(\Lambda_{11}\Lambda_{22}-\Lambda_{21}\Lambda_{12})=0, \tag{5.58}$$

in which Λ_{ij} are the elements of the matrix Λ. If all the matrix elements Λ_{ij} are real, as is the case for chemical systems, the solutions of the characteristic equation must be complex conjugate pairs. These solutions can be written as functions of the concentration [B]. Investigate whether their real parts (initially negative) can become positive due to an appropriate change in [B]. The point at which the real parts reach zero will be the point of transition from stability to instability.

For Eq. (5.58), since the coefficient of the linear term is the negative of the sum of the roots, if $\lambda_{\pm}$ are two roots, the following equation can be obtained:

$$\lambda_+ + \lambda_- = (\Lambda_{11} + \Lambda_{22}) = k_2[\mathrm{B}] - k_4 - k_3[\mathrm{X}]_\mathrm{S}^2. \tag{5.59}$$

If the real parts of this complex conjugate pair, $\lambda_{\pm}$, are negative then $k_2[\mathrm{B}] - k_4 - k_3[\mathrm{X}]_\mathrm{S}^2 < 0$; if they are positive, then $k_2[\mathrm{B}] - k_4 - k_3[\mathrm{X}]_\mathrm{S}^2 > 0$. Thus the condition that requires positive real parts for the onset of instability leads to

$$[\mathrm{B}] > \frac{k_4}{k_2} + \frac{k_3}{k_2}[\mathrm{X}]_\mathrm{S}^2 \text{ or } [\mathrm{B}] > \frac{k_4}{k_2} + \frac{k_3}{k_2}\frac{k_1^2}{k_4^2}[\mathrm{A}]^2. \tag{5.60}$$

Here Eq. (5.54) for $[\mathrm{X}]_\mathrm{S}$ had been used. Thus, for a fixed value of [A], as the value of [B] increases, when condition Eq. (5.58) is satisfied, the stationary state Eq. (5.54) becomes unstable, and this instability leads to chemical oscillations. Figure 5.8 illustrates theoretical results of chemical oscillations of concentrations [X] and [Y] on the basis of Brusselator or trimolecular model. Comparing with the oscillation curves in Fig. 5.2 of this chapter above, it can be found that this model is very successful in explanation of periodic changes.

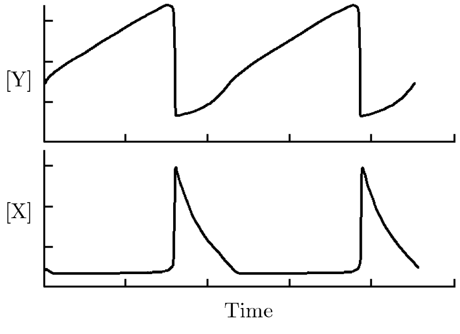


Fig. 5.8 Oscillations in [X] and [Y] of Brusselator

5.10 Turing Structures and Propagating Waves

In 1952, the British mathematician Alan Turing devised a simple model to show how chemical reactions and diffusion can work in consonance to produce stable stationary patterns of concentrations, which now is called Turing structures or stationary spatial dissipative structures.(p. 444)[5, 17, 18] In this way Turing tried to elucidate the possible mechanisms of biological morphogenesis and the formation of patterns in nature. Now it is known that biological morphogenesis is a very complex process, too complex to be explained entirely by the processes of chemical reactions and diffusion. However, since the

1970s Turing's observation has gained much attention due to the great interest in theoretical and experimental study of nonlinear nonequilibrium chemical systems. The Brusselator or trimolecular model of Sect. 5.9 will be used for discussion here.[3]

For simplicity, a system with one spatial dimensional coordinate r, in which diffusion occurs, will be under consideration, as shown in Fig. 5.9. (p. 446)[5] Supposing that the system extends from $-L$ to $+L$, the concentrations of reactants are maintained at a constant value at the boundaries, and the diffusion flows of the reactants are zero at the boundaries. Since diffusion flow is proportional to the derivative $\partial C/\partial r$, in which C is the concentration, the no-flow boundary conditions imply that the derivatives of the concentrations are zero at the boundary.

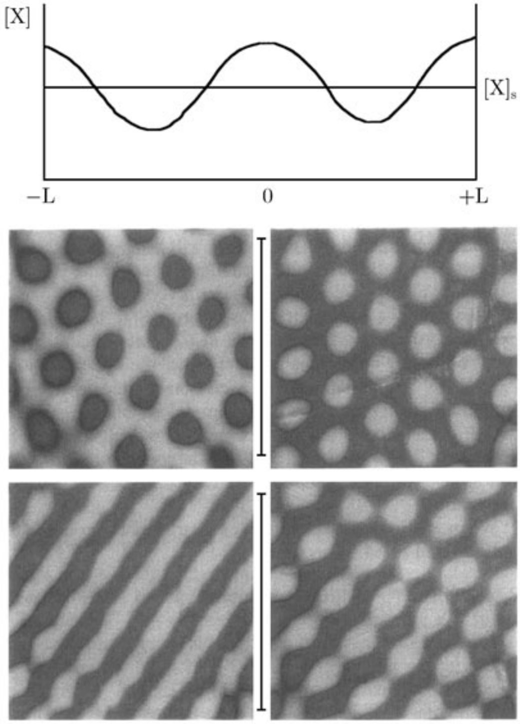


Fig. 5.9 (a) Turing structure in a one-dimensional Brusselator model, (b) One-dimensional Turing structure observed in chlorite-iodide-malonic acid reaction in an acidic aqueous solution (Courtesy Harry L. Swinney). The size of each square is nearly 1 mm. Reprinted from D. Kondepudi, I. Prigogine, *Modern Thermodynamics — From Heat Engines to Dissipative Structures* (John Wiley & Sons, New York, 1998) p. 446, Fig. 19.10 by permission of Wiley-Blackwell

When diffusion is included as a transport process, the kinetic Eq. (5.52) and Eq. (5.53) become

$$\frac{\mathrm{d}[\mathrm{X}]}{\mathrm{d}t} = D_{\mathrm{X}}\frac{\partial^2[\mathrm{X}]}{\partial r^2} + k_1[\mathrm{A}] - k_2[\mathrm{B}][\mathrm{X}] + k_3[\mathrm{X}]^2[\mathrm{Y}] - k_4[\mathrm{X}], \tag{5.61}$$

---

[3] This Section 5.10 is cited mainly from D. Kondepudi, I. Prigogine, *Modern Thermodynamics — From Heat Engines to Dissipative Structures* (John Wiley & Sons, New York, 1998) pp. 444–448

and

$$\frac{\mathrm{d}[\mathrm{Y}]}{\mathrm{d}t} = D_{\mathrm{Y}} \frac{\partial^2 [\mathrm{Y}]}{\partial r^2} + k_2 [\mathrm{B}][\mathrm{X}] - k_3 [\mathrm{X}]^2 [\mathrm{Y}]. \tag{5.62}$$

The boundary conditions are $\left.\frac{\partial[\mathrm{X}]}{\partial r}\right|_{r=-L} = \left.\frac{\partial[\mathrm{X}]}{\partial r}\right|_{r=+L} = 0$. In Eq. (5.61) and Eq. (5.62), D_{X} and D_{Y} are diffusion coefficients of X and Y, respectively; and r is the spatial coordinate. As before, it is assumed that the concentrations [A] and [B] are maintained at a fixed uniform value along the entire system (this assumption is introduced to simplify mathematics but it is difficult to achieve in practice). Diffusion usually homogenizes the concentration in a system, but when coupled with auto-catalytic chemical reactions under nonlinear nonequilibrium condition, it actually generates inhomogeneities or patterns. For pattern formation, the diffusion coefficients must be different. If the diffusion coefficients are nearly equal, then diffusion does not cause an instability; diffusion only tends to homogenize the instability that already exists. This can be seen as follows.

The stability of the stationary state Eq. (5.54) is first under consideration. The concentrations are homogeneous in the entire system. That is

$$[\mathrm{X}]_{\mathrm{S}} = \frac{k_1}{k_4}[\mathrm{A}] \text{ and } [\mathrm{Y}]_{\mathrm{S}} = \frac{k_2 k_4}{k_3 k_1}\frac{[\mathrm{B}]}{[\mathrm{A}]}. \tag{5.63}$$

The stability of this solution depends on the behavior of a small perturbation. If $\partial \mathrm{X}$ and $\partial \mathrm{Y}$ are the small perturbation from $[\mathrm{X}]_{\mathrm{S}}$ and $[\mathrm{Y}]_{\mathrm{S}}$, it is easy to see that equations linearized about the steady state Eq. (5.63) are

$$\frac{\partial}{\partial t}\begin{pmatrix}\delta \mathrm{X}\\ \delta \mathrm{Y}\end{pmatrix} = \begin{bmatrix} D_{\mathrm{X}}\frac{\partial^2}{\partial r^2} & 0 \\ 0 & D_{\mathrm{X}}\frac{\partial^2}{\partial r^2}\end{bmatrix}\begin{pmatrix}\delta \mathrm{X}\\ \delta \mathrm{Y}\end{pmatrix} + \begin{bmatrix} k_2[\mathrm{B}] - k_4 & k_3[\mathrm{X}]_{\mathrm{S}}^2 \\ -k_2[\mathrm{B}] & -k_3[\mathrm{X}]_{\mathrm{S}}^2 \end{bmatrix}\begin{pmatrix}\delta \mathrm{X}\\ \delta \mathrm{Y}\end{pmatrix}. \tag{5.64}$$

If it is assumed that $D_{\mathrm{X}} = D_{\mathrm{Y}} = D$, Eq. (5.64) can be written as

$$\frac{\partial}{\partial t}\begin{pmatrix}\delta \mathrm{X}\\ \delta \mathrm{Y}\end{pmatrix} = D\frac{\partial^2}{\partial r^2} I \begin{pmatrix}\delta \mathrm{X}\\ \delta \mathrm{Y}\end{pmatrix} + M\begin{pmatrix}\delta \mathrm{X}\\ \delta \mathrm{Y}\end{pmatrix}. \tag{5.65}$$

Here M is the matrix in the second term of Eq. (5.64) and I is the identity matrix. For a linear equation of this type, the spatial part of solution can always be written as combination of sin Kr and cos Kr, in which wavenumbers K are chosen so that the boundary conditions are satisfied. This means that if the behavior of the spatial part of a perturbation is separated,

$$\begin{pmatrix}\delta \mathrm{X}(t)\\ \delta \mathrm{Y}(t)\end{pmatrix} \sin Kr \text{ and } \begin{pmatrix}\delta \mathrm{X}(t)\\ \delta \mathrm{Y}(t)\end{pmatrix} \cos Kr, \tag{5.66}$$

then the behavior of all linear combinations of these basic solutions can be deduced. If Eq. (5.66) is substituted into Eq. (5.65), then

$$\frac{\partial}{\partial t}\begin{pmatrix}\delta \mathrm{X}\\ \delta \mathrm{Y}\end{pmatrix} = \left(-DK^2 I + M\right)\begin{pmatrix}\delta \mathrm{X}\\ \delta \mathrm{Y}\end{pmatrix} \tag{5.67}$$

is obtained. From this expression it is clear that if λ_+ and λ_- are the eigenvalues of M, the addition of diffusion will only change the eigenvalues to $(\lambda_+ - DK^2)$ and $(\lambda_- - DK^2)$. Since it is the positivity of the real part of the eigenvalue that indicates instability, it can be seen that (in this case, where $D_X = D_Y$), diffusion does not generate a new instability; it only makes steady states more stable to perturbations with $K \neq 0$. Therefore, the solution to Eq. (5.67) with $K = 0$ is the least stable state because its eigenvalues will have the largest real part.

For the emergence of spatial patterns, the diffusion coefficients must be unequal. In a small region, if one species diffuses out more rapidly than the other, the growth of one species may be facilitated by the depletion of the other. If this happens the homogeneous state will no longer be stable and inhomogeneities will begin to grow. When the diffusion coefficients are unequal it is easy to see that in place of the matrix $(-DK^2 I + M)$, the following matrix may be obtained.

$$\begin{bmatrix} k_2[\mathrm{B}] - k_4 - K^2 D_X & k_3[\mathrm{X}]_{\mathrm{S}}^2 \\ -k_2[\mathrm{B}] & -k_3[\mathrm{X}]_{\mathrm{S}}^2 - K^2 D_X \end{bmatrix}. \tag{5.68}$$

For an instability to produce stationary spatial structure, the two eigenvalues of this matrix must be real, and at least one must become positive. If these eigenvalues are real, and one becomes positive due to the variations in the parameters [B] and [A], then the unstable perturbation will be of the following form.

$$\begin{pmatrix}c_1\\ c_2\end{pmatrix}\sin(Kr)e^{\lambda_+ t} \text{ and } \begin{pmatrix}c_1\\ c_2\end{pmatrix}\cos(Kr)e^{\lambda_+ t}e^{\lambda_+ t}, \tag{5.69}$$

in which λ_+ is the eigenvalue with positive real part. This indicates a growth of a spatial pattern $\sin Kr$ or $\cos Kr$ without any temporal oscillations; it will evolve to a stationary pattern or a Turing structure.

On the other hand, if the eigenvalues are a complex-conjugate pair, then the solutions to the perturbation Eq. (5.64) will be of the form

$$\begin{pmatrix}c_1\\ c_2\end{pmatrix}\sin(Kr)e^{(\lambda_{\mathrm{re}}+i\lambda_{\mathrm{im}})t} \text{ or } \begin{pmatrix}c_1\\ c_2\end{pmatrix}\cos(Kr)e^{\lambda_+ t}e^{(\lambda_{\mathrm{re}}+i\lambda_{\mathrm{im}})t}, \tag{5.70}$$

in which $\begin{pmatrix}c_1\\ c_2\end{pmatrix}$ is the eigenvector with the eigenvalue $\lambda = \lambda_{\mathrm{re}} + i\lambda_{\mathrm{im}}$, with its real and imaginary parts as shown. If the real part λ_{re} is positive, the perturbation Eq. (5.70) will grow. The unstable perturbation contains oscillations in time, due to the factor $e^{i\lambda_{im}t}$, as well as variations in space due to the factor sin Kr or cos Kr. Such a perturbation corresponds to a propagating wave.

From a simple kinetic viewpoint,[4] B-Z reactions are auto-catalytic reduction-oxidation (redox) processes. Each step of the multi-step reaction sequence generates a catalyst that

---

[4] http://www.meta-synthesis.com/webbook/24_complexity/complexity3.html. Cited 1 Dec 2008

speeds up the counter reaction. As the reduction step proceeds it produces more and more catalyst to speed up the oxidation, and vice versa. In some B-Z reactions, the reduced state may be red while the oxidised state may be a pale brown. If the reaction system is not stirred, diffusion control takes over. This can be observed when the reaction mixture is prepared as a thin film, 0.5 – 1 mm deep, to avoid any vertical mixing, and waves of reaction can be seen propagating through the unstirred mixture.

Experimentally, travelling waves have been observed in the B-Z reaction, as shown in Fig. 5.10, but only recently have the Turing patterns been observed in the laboratory. (p. 450)[5]

Fig. 5.10 Traveling waves in the B-Z reaction. Reprinted from http://http://www.intute.ac.uk/sciences/spotlight/issue6/waves.html by permission of Mark R. Leach of Meta-Synthesis, UK. Cited on 1 Dec 2008

5.11 Prigogine's Carelessness on Thermodynamic Coupling

Prigogine's researches on nonlinear nonequilibrium thermodynamics should belong to nonlinear dissipative thermodynamics. However, out-of-thermodynamics nonlinear kinetic methods and approaches were mainly accepted by Prigogine and his coworkers, so Prigogine dissipative structures and related work were out of thermodynamics course contents in many thermodynamics textbooks.

As mentioned above, Prigogine agreed with the concept of thermodynamic coupling, which had been firstly and clearly proposed by a biochemist D. Burk. The following Prigogine's words,

> "Thermodynamic coupling allows one of the reactions to progress in a direction contrary to that prescribed by its own affinity."

can be regarded as the definition of thermodynamic coupling. However, just due to the complexity of biochemical reactions and that no really coupled reaction examples were found in ordinary chemistry at that time, a so-called "under standard conditions" was incorrectly introduced by the Brussels thermodynamics school. In this way, the correct mathematical expression of thermodynamic coupling [$\mathrm{d_i}S_1 < 0$, $\mathrm{d_i}S_2 > 0$ & $\mathrm{d_i}S \geqslant 0$] was practically changed into an incorrect one [$\mathrm{d_i}S_1^0 < 0$, $\mathrm{d_i}S_2^0 > 0$ & $\mathrm{d_i}S^0 \geqslant 0$], so-called "under standard conditions". In particular, after Prigogine won the 1977 Nobel Prize, such a misleading concept spread quickly with Prigogine's reputation. Even in his last years, in the book "*Modern Thermodynamics - From Heat Engines to Dissipative Structures*" written by Kondepudi and Prigogine, an ordinary chain reaction of HBr formation from Br_2 and H_2 was regarded as an example of thermodynamic coupling. The root of incorrectness was that the so-called criterion of "under standard conditions"

$[(\Delta G_1^0)_{T,p} > 0,\ (\Delta G_2^0)_{T,p} < 0\ \&\ (\Delta G^0)_{T,p} \leqslant 0]$ was accepted by them. This issue has been discussed in detail in Chapter 4, and so no more is needed here.

5.12 Thermodynamic Coupling Model of Spiral Reactions

The concept of thermodynamic coupling is the core of modern thermodynamics, and modern thermodynamics is thermodynamics for researches on complex (coupling) systems. Now, the well-known B-Z reactions will be taken as a practical example of thermodynamic coupling for further discussion in the field of dissipative thermodynamics.

The kinetic mechanisms of B-Z reactions are usually very complex. To increase our grasp on the matter, it is better to do an experimental demonstration of B-Z reactions, or to see a video of the experimental demonstration. It will especially be much better for teachers and students of physics disciplines if even discussions on chemistry and chemical reactions are worrisome for them. In fact, the experiments of B-Z reactions are not too difficult.[6] In a beaker with a stirrer, dissolve 4.292 g malonic acid [$CH_2(COOH)_2$] and 0.175 g ammonium cerium (IV) nitrate [$Ce(NH_4)_2(NO_3)_6$] into 150 ml 1 M sulfuric acid. After a short time the solution is clear, and 1.415 g sodium bromate ($NaBrO_3$) is added. After a few minutes, the color of the solution will oscillate between yellow and colorless. The period of oscillation is about 1 to 2 minutes. Such a phenomenon is called chemical oscillation. Figure 5.2 above showed that the concentrations of ions changed periodically. (Caution: 1. Experiments should be made under guidance of a chemistry teacher. 2. If concentrated sulfuric acid should be diluted into 1 M sulfuric acid, then the concentrated sulfuric acid should be slowly injected into water whilst stirring. Never inject water into the concentrated sulfuric acid otherwise an explosion or splash may take place. In case one's body comes into contact with sulfuric acid, then rinse with a large amount of water).

The well-accepted mechanism was proposed by Field, Körös and Noyes, so the mechanism was called the Field-Körös-Noyes mechanism or simply the FKN mechanism. (p. 257)[6] The FKN mechanism consists of about 20 steps. These steps have been summarized into the following three main reaction chains. Those are reaction chain A, reaction chain B and reaction chain C taking place in different intervals of time. The total effects of each reaction chain are as follows.

$$\begin{aligned}\mathrm{A:}\ \mathrm{BrO_3^-} + 2\,\mathrm{Br^-} &+ 3\,\mathrm{CH_2(COOH)_2} + 3\,\mathrm{H^+} \\ &= 3\,\mathrm{BrCH(COOH)_2} + 3\,\mathrm{H_2O},\end{aligned} \tag{5.71}$$

$$\mathrm{B:}\ \mathrm{BrO_3^-} + 4\,\mathrm{Ce^{3+}} + 5\,\mathrm{H^+} = \mathrm{HOBr} + 4\,\mathrm{Ce^{4+}} + 2\,\mathrm{H_2O}, \tag{5.72}$$

$$\begin{aligned}\mathrm{C:}\ \mathrm{HOBr} + 4\,\mathrm{Ce^{4+}} &+ \mathrm{BrCH(COOH)_2} + \mathrm{H_2O} \\ &= 2\,\mathrm{Br^-} + 4\mathrm{Ce^{3+}} + 3\,\mathrm{CO_2} + 6\,\mathrm{H^+}.\end{aligned} \tag{5.73}$$

From the viewpoint of classical thermodynamics, any isothermal isobaric chemical oscillation would consist of a forward reaction with decreasing Gibbs free energy and a backward reaction with increasing Gibbs free energy in different intervals of time alternately, so chemical oscillations had been disbelieved and regarded as thermodynamically impossible. However, from the viewpoint of modern thermodynamics, i.e. from viewpoint of thermodynamic coupling, chemical oscillation can be satisfactorily explained.[1,8,19~22]

1. If the chemical oscillation is running under constant temperature and constant pressure, from the mathematical expression of the second law of thermodynamics it is known that all Gibbs free energies of reaction chain A, reaction chain B and reaction chain C taking place in three different intervals of time must decrease, i.e. $\Delta G_{\mathrm{A}} < 0$, $\Delta G_{\mathrm{B}} < 0$ and $\Delta G_{\mathrm{C}} < 0$.
2. Reaction chain A is a spontaneous reaction, so it is normal that reaction chain A is running toward the right side. Reaction chain B is also a spontaneous reaction, so it is also normal that reaction chain B is running toward the right side.
3. Now reaction chain C might be a key among them. Chemical oscillation requires that the reaction chain C might include a reversed reaction chain $(-\mathrm{A})$ and a reversed reaction chain $(-\mathrm{B})$ simultaneously, i.e. $\mathrm{C} = [(-\mathrm{A}) + (-\mathrm{B}) + \mathrm{W}]$. Of course, the reversed reaction chain $(-\mathrm{A})$ and the reversed reaction chain $(-\mathrm{B})$ are nonspontaneous reactions, i.e. $\Delta G_{-\mathrm{A}} > 0$ and $\Delta G_{-\mathrm{B}} > 0$. Based on the criterion of thermodynamic coupling $[\Delta G_{[(-\mathrm{A})+(-\mathrm{B})]} > 0$, $\Delta G_{\mathrm{W}} < 0$ & $\Delta G_{\mathrm{C}} < 0]$, the reaction W must be a very strong spontaneous and energy-providing reaction. That is true for the reaction W. The whole reaction $\mathrm{W} = (\mathrm{A} + \mathrm{B} + \mathrm{C})$

$$\begin{aligned}\mathrm{W}:\ 2\,\mathrm{BrO_3^-} &+ 3\,\mathrm{CH_2(COOH)_2} + 2\,\mathrm{H^+} \\ &= 2\,\mathrm{BrCH(COOH)_2} + 3\,\mathrm{CO_2} + \mathrm{H_2O}.\end{aligned} \tag{5.74}$$

   is an oxidation of an organic compound, which, of course, is an energy-providing or energy-consumption reaction. Due to the energy provided from the reaction W, the reversed reaction chain $(-\mathrm{A})$ and the reversed reaction chain $(-\mathrm{B})$ can be realized, and the whole reaction system is basically back to the start point of the next period of the chemical oscillation. In the whole reaction W malonic acid $[\mathrm{CH_2(COOH)_2}]$ is oxidized by a strong oxidant, bromate ions ($\mathrm{BrO_3^-}$ is formed from sodium bromate and sulfuric acid) in an acidic circumstance, and partially changed into carbon dioxide and water, which have nearly no influences on the reaction system left. Therefore, after a reaction period all intermediate ions in the solution have been recovered into their original concentrations. In particular, the valence state of cerium ions as a catalyst has recovered into its original state (the color of cerium ions in valence state IV is yellow. The color of cerium ions in valence state III is colorless. After recovering, the valence state of cerium ions changes back with yellow color again), so the second or next period of chemical oscillation starts.

Note that the reaction W is the sum of the reaction chain A, the reaction chain B and the reaction chain C, i.e. $\mathrm{W} = (\mathrm{A} + \mathrm{B} + \mathrm{C})$. The total result is the consumption of a small part of bromate ions ($\mathrm{BrO_3^-}$ is formed from sodium bromate and sulfuric acid) and organic malonic acid $[\mathrm{CH_2(COOH)_2}]$. Byproducts are bromomalonic acid $[\mathrm{CHBr(COOH)_2}]$, carbon dioxide $\mathrm{CO_2}$ (to form gas bubble and leave from the solution) and water $\mathrm{H_2O}$ (remaining in solution). All of these by-products have nearly no obvious influences on the next period of oscillation, so the chemical oscillation usually lasts a long time (about a half hour) until the main reactants, malonic acid [$CH_2(COOH)_2$] and sodium bromate ($NaBrO_3$), are depleted. The observable color oscillation is just the color of different valance states of cerium ions. The apparatus of such an experimental demonstration of B-Z reactions is very simple and has a low cost for understanding modern thermodynamics. B-Z reactions can also be used to show macroscopic ordering patterns.

The explanation from modern thermodynamics above for chemical oscillations seems to be satisfied, but in 2005 the author undertook a further consideration about chemical oscillation, and got some new findings. From the equation W = (A + B + C), three written styles of A = [(−B) + (−C) + W], B = [(−C) +(−A) + W] and C = [(−A) + (−B) + W] can be obtained. Up to now, only the third written style, in which the reaction chain C seems to be the key reaction in thermodynamic coupling, has been discussed. Based on the first or the second written styles, in which the reaction chain A or the reaction chain B seems to be the key reaction in thermodynamic coupling, similar results can be obtained. It means that the reaction chain A, the reaction chain B and the reaction chain C play similar important roles in chemical oscillation, and that this chemical oscillation does not consist of a simple pair of a forward reaction and a backward reaction, but consists of the reaction chain A, the reaction chain B and the reaction chain C as an approximately cyclical reaction. It is more accurate to say that this chemical oscillation is a dissipative spiral reaction, as shown in Fig. 5.11. The experimentally observed phenomenon of chemical oscillation, as shown in Fig. 5.2, is only a view from one side or two sides of the complete dissipative spiral reaction. The cyclical reaction, as shown in Fig. 5.4, is a planform (the view from up to down) of the dissipative spiral reaction. The reaction W provides nearly the same promotion via thermodynamic coupling to the reaction chain A, the reaction chain B and the reaction chain C. After each period of the spiral reaction, the Gibbs free energy level goes down a little due to the dissipation. Until the organic acid and the oxidant for providing energy are exhausted, the chemical oscillation stops.

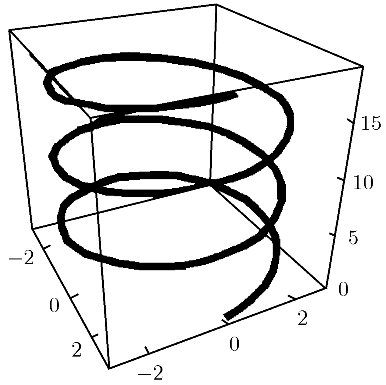


Fig. 5.11 Spiral reactions

The spiral reactions, i.e. the nature of B-Z reactions, was firstly reported by the author in 2005.[1] No report about spiral reactions had been found before this. The spiral reactions provide many useful revelations for thermodynamics.

1. To discover spiral reactions there is no necessity to introduce any out-of-thermodynamics assumption. It is directly based on the second law of thermodynamics for complex (coupling) systems, i.e. the mathematical expression of thermodynamic coupling. Therefore, the author calls it thermodynamic coupling model of spiral reactions, or thermodynamic coupling model for chemical oscillations. Of course, the treatment for nonequilibrium states must include the local equilibrium approximation.
2. After knowing the nature of B-Z reactions, chemical oscillations are much easier to understand. In mechanics, if there is a small ball with a horizontal initial speed inside

a cylindrical container, then due to the inertia and gravity its moving locus will follow a spiral curve going down. For an isothermal isobaric B-Z reaction, the axes of the spiral curve may be a coordinate of Gibbs free energy. During the progress of reaction Gibbs free energy of a complex (coupling) system may be going down along a spiral curve. That is completely in agreement with the prediction of the second law of thermodynamics: Gibbs free energy of an isothermal isobaric practical reaction system must continuously decrease. Generally, accumulated value of entropy production of B-Z reaction should be taken as the axes coordinate of the spiral curve. During the progress of the B-Z reaction, its accumulated value of entropy production continuously increases to form a spiral curve.

3. Looking from up to down the spiral reaction, it shows a cyclical reaction. Usually, cyclical reactions are regarded as belonging to linear dissipative thermodynamics. As discussed above, from cyclical reaction Onsager reciprocal relations had been deduced. However, the side views of spiral reactions are chemical oscillations. Usually, chemical oscillations are regarded as a typical example of Prigogine dissipative structures, which belong to nonlinear dissipative thermodynamics. Therefore, there is no obvious break between linear dissipative thermodynamics and nonlinear dissipative thermodynamics, so there should be no obvious gap for corresponding thermodynamic rules between them.
4. Spiral reactions also show that the complexity of B-Z reaction systems provides material basis to realize dissipation decrease theorem. During Gibbs free energy decline (isothermal isobaric systems) or entropy production (any systems), thermodynamic coupling chooses a near-horizontal spiral-like line to decrease dissipation. If the container of a B-Z reaction is completely isolated from exterior as an isolated system, then entropy value of the system will increase continuously along a spiral curve, and finally researches its maximum. As mentioned above, from a chemistry viewpoint, B-Z reaction is an oxidation-reduction reaction of W. It seems to be completed in one step, but it is practically completed in A, B, C, some reaction chains along a cyclical reaction. In the macroscopic aspect it finishes a certain of functions (such as color changes). These reaction chains of the B-Z reaction represent the natural rule of dissipation decrease theorem.
5. The second law of thermodynamics includes two parts of equality and inequality. In modern thermodynamics, the equality part is the ideal situation of nondissipation, i.e. the spiral reaction becomes a plane single-way cyclical reaction. Therefore, in his famous paper, Onsager also pointed out that

   > "we should have addition transitions taking place around the cycle",

   i.e. the single-way cycle, in his famous paper. However, from the single-way cycle, Onsager reciprocal relations would not be obtained.
6. Obviously, if an open system can be constructed with a continuous supplement of dissipated materials (organic acid and oxidant, here) and with a continuous removal of the reaction products, then the reaction will be a real plane single-way cyclical reaction. In metabolism of living systems, the jumpiness of the heart may be regarded as an example of such nearly ideal single-way cyclical processes. This periodic chemical oscillation of an inanimate system opens wide thought in research on the biological clock of liv-

ing systems. Modern thermodynamics will strongly promote the development of living sciences and advanced technology in the future.

References

1. Wang J-T. Modern thermodynamics – and a whole view of thermodynamics [M]. Shanghai: Fudan Univ. Press, 2005, 223 pages.
2. Wang J-T, Zheng P-J. Extended Carnot Theorem. 19th Internat. Conf. on Chem. Thermodynamics [C]. Boulder, CO, USA: 30 Jul - 4 Aug 2006, http://www.symp16.nist.gov/pdf/p745.pdf. Cited 20 Apr 2008.
3. Holman J P. Thermodynamics [M]. 3rd edn. New York: McGraw-Hill, 1980, 1.
4. Onsager L. Physical Review [J], 1931, 37: 405; 1931, 38: 2265.
5. Kondepudi D, Prigogine I. Modern Thermodynamics — From Heat Engines to Dissipative Structures [M]. New York: John Wiley & Sons, 1998. http://www.wiley.com/
6. Li R-S. Nonequilibrium Thermodynamics and Dissipative Structure [M], Beijing: Tsinghua University Press, 1986.
7. Deng J-F, Fan K-N. Physical Chemistry [M]. Beijing: High Education Press, 1993.
8. Wang J-T. Nonequilibrium Nondissipative Thermodynamics – With Application to Low-Pressure Diamond Synthesis [M]. Heidelberg: Springer, 2002.
9. Prigogine I. Acad Roy Belg, Bull Clesse Sci [J], 1945, 31: 600.
10. Prigogine I. Introduction to Thermodynamics of Irreversible Processes [M]. 3rd edn. New York: Interscience Publishers, John Wiley & Sons, 1967, 76.
11. Thomson W (Lord Kelvin). Proc Roy Soc [C]. Edinburgh: 1854, 123; Collected Papers I, 237-242.
12. Von Helmholtz H. Wied Ann [J], 1876, 3: 201; Wiss Abh [J], 1: 840.
13. Nernst W. Zeits F physik Chem [J], 1888, 2: 613.
14. Chandrasekhar S. Hydrodynamic and Hydromagnetic Stability [M]. 3rd press. New York: Dover Publication Inc, 1981; Oxford: Clarendon Press, 1961, Chap II, 10-12.
15. Mayer-Spasche R. Pattern Formation in Viscous Flows [M]. Switzerland: Birkhauser Verlag, 1999, Fig.4.1 on page 149; van Dyke M D. An Album of Fluid Motion [M], Stanford: The Parabolic Press, 1982, Fig.142; Original from Koschmieder EL Adv Chem Phys [J], 1974, 26: 177.
16. Shen X-F, Hu G, Jiang L. Theory of Dissipative Structures [M]. Shanghai: Shanghai People's Publisher, 1987, 36-45.
17. Turing A. Phil. Trans. R. Soc. London [J], 1952, B237: 37; Kondepudi D, Prigogine I. Modern Thermodynamics — From Heat Engines to Dissipative Structures [M]. New York: John Wiley & Sons, 1998, 444.
18. Kapral R, Showalter K (eds). Chemical Waves and Patterns [M]. New York: Kluwer, 1994.
19. Wang J-T, Zheng P-J. The Second Law in Modern Thermodynamics. 19th Internat Conf on Chem Thermodynamics [C]. Boulder, CO, USA: 30 Jul - 4 Aug 2006, http://www.symp16. nist.gov/pdf/p2086.pdf. Cited 20 Apr 2008.
20. Wang J-T. Modern Thermodynamics in the 21st Century (plenary lecture). 2006 National Symp on Thermodynamics and Statistics [C]. Lanzhou, China: 17 Jul 2006.

21. Wang J-T. Modern Thermodynamics of the 21st Century and Nonequilibrium Phase Diagrams for Carat-size Synthetic Diamond. Proc. of the 13th National Symp. (Chinese and Japanese Join Meeting) on Phase Diagrams [C]. Xiamen, China: 9 Nov 2006, 245.
22. Wang J-T, The Second Law of Thermodynamics in the Current 21st Century (plenary lecture). 2007 National Symp on Thermodynamics and Statistics [C]. Yanji, China: 6 Aug 2007.

Chapter 6

Thermodynamics Coupling Model for Activated Low-Pressure Diamond Growth

Abstract There is a deep historical relationship between thermodynamics and diamond synthesis. The success of high-pressure diamond synthesis can be regarded as a brilliant achievement under the theoretical guide of classical thermodynamics. Current nondissipative thermodynamics in modern thermodynamics are the direct result of the need to explain the new activated low-pressure diamond synthesis. The activated low-pressure diamond synthesis was first recognized by former Soviet scientists in about 1970, and made fun of as “alchemy” by some Western scientists at that time. Later, due to Japanese scientists’ work the new diamond technique spread into the West, and formed a research upsurge in the world in the latter 1980s and up to now. However, the thermodynamic explanation for the new technology is still based on our thermodynamic coupling model. An introduction of the thermodynamic coupling model and the problems and bepuzzlements associated with all other classical thermodynamics models will be discussed in this chapter. Note that any practical macroscopic processes including the new diamond process belong to dissipative thermodynamics, but all calculations of nonequilibrium phase diagrams for the low-pressure diamond synthesis belong to nondissipative thermodynamics, which will be mainly discussed in the following chapters, and this Chapter will be a link or transition from dissipative thermodynamics to nondissipative thermodynamics.

6.1 High-Pressure Diamond Syntheses

It has been said that natural diamond was found probably several thousand years ago. The first report about diamond was written in the Bible, which mentioned it existing in the 12th century B.C. Diamond has always seemed to be a kind of rare precious treasure or a symbol of wealth. Diamond is the hardest material. The English name “diamond” originated from the Greek word “αδαμαζ” which means unsubjugable, so the character of diamond can be seen. Moreover, diamond has many other outstanding characteristics. For instance, the heat conductivity of diamond is better than that of silver and copper. Diamond is economically very valuable. In the 17th or 18th centuries, investigation started to explore the composition of diamond. For instance, in 1675 Isaac Newton (1641–1727), on the basis of comparison between refractive index of diamond and that of some combustible gases, pointed out that diamond must also be combustible. After that, Robert Boyle (1627–1691) confirmed that diamond underwent much change in the flame. In 1694, the Italian Academy of Science organized a performance to show diamond burning out at the focus of a lens. In this way Newton's judgment had been demonstrated. In 1772, A.L. Lavoisier (1743–1794) studied the gas formed by combustion of diamond. It was found that the gas made lime-water turbid. Therefore, it was known that carbon dioxide was produced. In 1797, S. Tennant burnt diamond in a closed metallic container filled with oxygen. The carbon content of the produced gas was just the same as the weight of the burnt diamond. In this way, it was demonstrated that diamond is a crystal form of carbon. Later in 1799, H. Guilton demonstrated that graphite is another crystal form of carbon, so graphite and diamond are allotropes. Coal is also an allotrope of carbon, but it is an amorphous form of carbon with more impurities. The price difference between diamond and graphite is so large that the synthesis of diamond from graphite (or coal) is very attractive. [1]

Given that diamond has the highest density among all the allotropes of carbon and a close relationship exists between the original ore of natural diamond and volcanic kimberlite (named after the town Kimberley in South Africa and famous for its role of diamond carrier), it was supposed that the formation of synthetic diamond should be under high temperature and high pressure. Air should also be isolated from the produced diamond to prevent it from burning. In the 19th century and at the beginning of the 20th century the experimental conditions were very simple and crude. Soldered sealed iron tubes including graphite were usually put into a furnace to produce the imaginary "high temperature and high pressure" conditions. The success of synthetic diamond or diamond-like crystals had been declared several times. However, based on experimental repetition none of these reports could be confirmed. In 1928, a paper entitled "*The Problem of the Artificial Production of Diamond*" was published in *Nature*. After a careful analysis of the existing evidence, the writer stated: [2]

> The conclusion seems inevitable that diamonds have not yet been produced in the laboratory, and that investigators have been misled into regarding as diamond various transparent, single-refracting minerals which happen to be very resistant to chemical reagents.

In 1938, F.D. Rossini and R.S. Jessup calculated the equilibrium line between the diamond and graphite phases below 1200 K according to data of the combustion heat, specific heat, compressibility, and heat expandability of both diamond and graphite. [3] On this basis, R. Berman and F. Simon made some revisions, taking the change of the Gibbs free energy change ΔG as a criterion of phase equilibrium, and extended the phase line between diamond and graphite to higher than 1200 K. [4] That is the Berman-Simon line, which is well known in the forum of diamond synthesis. Its mathematical expression is

$$\begin{aligned}\Delta G_1\ (\mathrm{kJ\ mol^{-1}}) = \{&p\ (\mathrm{Pa}) - [2.73\times10^6 T\ (\mathrm{K}) \\ &+7.23\times10^8]\}\Delta V\ (\mathrm{m^3\ mol^{-1}}).\end{aligned} \tag{6.1}$$

To help understand the predictability and guidance that thermodynamics offers, the following transformation reaction from graphite, denoted by C (gra), into diamond, denoted by C (dia), could be taken as an example to illustrate a simple calculation and discussion. [5, 6]

$$\mathrm{C(gra)} = \mathrm{C(dia)}. \tag{6.2}$$

From thermodynamic databooks, the heat of formation of C (gra) and C (dia) at standard temperature (298.15 K) and standard pressure (1 atmosphere, p^0 = 101.325 kPa) are 0 and

1.8962 kJ mol^{-1}, respectively. The entropies of C (gra) and C (dia) at standard temperature and standard pressure are 5.6940 kJ K^{-1} mol^{-1} and 2.4389 kJ K^{-1} mol^{-1}, respectively. By calculation, it can be shown that

$$\Delta H^0_{298} = 1.8962 - 0 = 1.8962\ \text{kJ mol}^{-1}, \tag{6.3}$$

$$\Delta S^0_{298} = 2.4389 - 5.6940 = -3.2551\ \text{J K}^{-1}\text{mol}^{-1} \tag{6.4}$$

and

$$\begin{aligned}\Delta G^0_{298} &= \Delta H^0_{298} - T\Delta S^0_{298} \\ &= 1.8962 - (298)\times(-0.0032551) = 2.866\ \text{kJ mol}^{-1}.\end{aligned} \tag{6.5}$$

That is, the formation Gibbs free energies of C (gra) and C (dia) at standard temperature and standard pressure ΔG^0_{298} are 0 and 2.87 kJ mol^{-1}, respectively. It is also known that the densities of C (gra) and C (dia) at standard temperature and standard pressure are 2.261×10^3 and 3.513×10^3 kg m^{-3}, respectively. From Eq. (3.22) in Chapter 3 Fundamentals of Classical Thermodynamics it is known that

$$\left(\frac{\partial G}{\partial p}\right)_T = V. \tag{6.6}$$

The corresponding equation can be obtained for a reaction as

$$\left(\frac{\partial \Delta G}{\partial p}\right)_T = \Delta V. \tag{6.7}$$

Therefore, it may be possible to change the pressure to get a negative value of the Gibbs free energy of the transformation from graphite to diamond.

$$\Delta G(p) - \Delta G\left(p^0\right) = \int_{p^0}^{p} \Delta V \mathrm{d}p = \Delta V\left(p - p^0\right), \tag{6.8}$$

$$\begin{aligned}\Delta G(p) &= \Delta G\left(p^0\right) + \Delta V\left(p - p^0\right) = 2.87\ \text{kJ mol}^{-1} \\ &+ \left[\left(\frac{12.011}{3.513} - \frac{12.011}{2.260}\right)\times 10^{-6}\text{m}^3\text{mol}^{-1}\right]\times(p - 101.325\ \text{kPa}).\end{aligned} \tag{6.9}$$

If $\Delta G(p) < 0$ was required, then the pressure should be $p > 1.52\times10^9$ Pa, i.e. about 15 000 times that of the atmosphere. It was predicted that graphite might be transformed into diamond at very high pressure.

Based on the theoretical prediction of classical thermodynamics and the technical support of high-pressure physics, on 16 December 1954 F.P. Bundy, H.T. Hall and their coworkers at the American General Electric Company were successful with high pressure processes, obtaining the first synthetic diamond crystals from graphite or other carbon materials. In 1955, Bundy and Hall et al. announced the achievement of high pressure diamond synthesis in the journal *Nature*.[7] Metals were added as catalysts into the reaction mixture. That was very important for the success of the process, so the process was called a high-pressure high-temperature catalyst process (HPHT catalyst process), or simply a high-pressure catalyst process (HP catalyst process). Now it is estimated that if there were no catalysts, the transformation from graphite to diamond would be achieved at about 15 GPa and 3000–4000 K. Up to now, no available container could withstand these conditions. If metallic catalysts (nickel, molybdenum, cobalt and so on) are adopted, synthetic diamond can be obtained at 5.5–6 GPa and 1600–1800 K in about 1 minute. Figure 6.1 shows a freshly grown single crystal of diamond, half-buried in the center of its catalyst metal growth medium, now frozen.[8, 9] It was prepared at 6 GPa and 1700–1800 K by using nickel as the catalyst. Later on, after continuous improvements the high pressure catalyst process had become the main process of diamond syntheses in industry. The production of synthetic diamonds by the high pressure catalyst process worldwide is about 600 million carats (120 thousand tons) per year. Synthetic diamonds are widely used in different industrial applications. At the same time, the success of high pressure diamond synthesis has also confirmed the importance of thermodynamics – one of the fundamental scientific disciplines.

After 15 years of effort, in 1970 the American General Electric Company announced the success of high-pressure synthetic gem-quality diamond and it was said that several transparent diamonds with 0.30, 0.31 and 0.39 carat, respectively, were obtained. However, the production cost was too high, so it was still not able to compete with natural diamond. Recent reports in the 21st century say that carat-size gem-quality diamonds have been grown by low-pressure chemical vapor deposition (CVD) process, which will be discussed in the last chapter of the book.

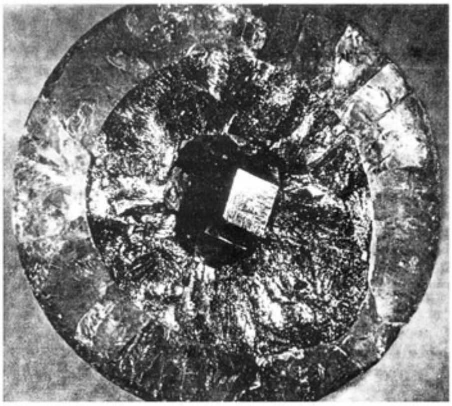


Fig. 6.1 A single crystal of diamond prepared by high-pressure catalyst process. Reprinted from W.J. Moore, *Physical Chemistry*, 4th edn. (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1972) Chap. 6, p. 226, and originally from R. H. Wentorl, *General Electric R & D Center*, New York (1970). Details were reported at the Symposium on Rate Processes at High Pressure, Toronto, May 1970

6.2 Activated Low-Pressure Diamond Growth from the Vapor Phase

B.V. Deryagin, B.V. Spitsyn, and D.V. Fedoseev et al. of the Institute of Physical Chemistry in former Soviet Union Moscow made a sustained research effort on the activated low-pressure diamond growth, and in about 1970 they met with success, getting very beautiful crystals of diamond on nondiamond substrates. [10~16] Photographs of their low-pressure synthetic diamond crystals were published in 1976, as shown in Fig. 6.2. However, their work was not well understood in the West, and nearly nobody believed the achievement of the low-pressure diamond synthesis. Moreover, because of the spurious or mistaken claims for the success of diamond synthesis repeatedly made in the 19th century and in the beginning of the 20th century, together with the lack of high-pressure equipment, the diamond synthesis under low pressures was regarded as impossible. A further impression was gained from classical equilibrium thermodynamics. This prediction or conclusion came only from equilibrium premise. The activated low-pressure diamond growth from the vapor phase was still considered to be absolutely thermodynamically impossible.

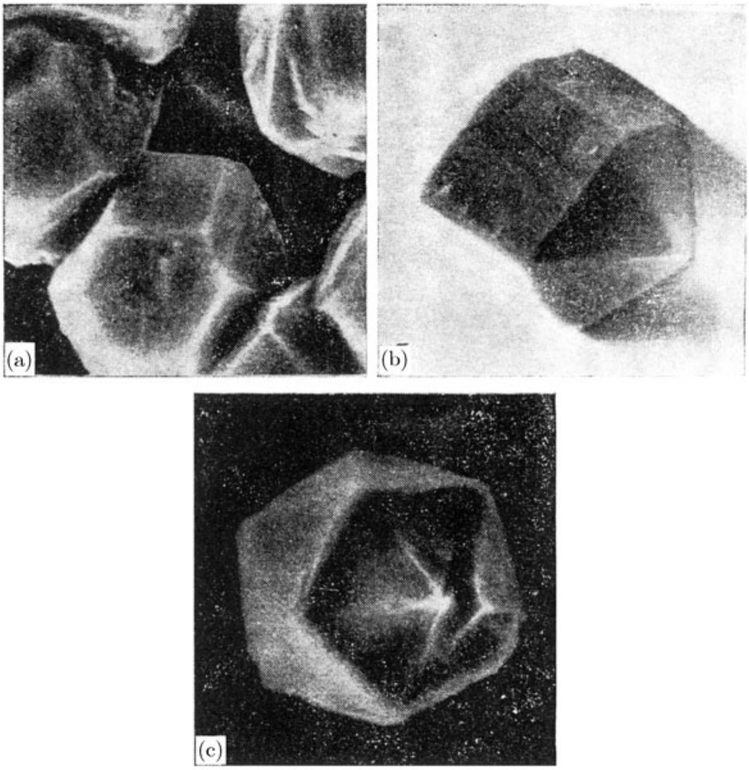


Fig. 6.2 Beautiful diamond crystals grown from the vapor phase on a nondiamond substrate by Deryagin, Spitsyn et al. in 1976. [12] Reprinted from B.V. Deryagin, B.V. Spitsyn et al. *Dokl. Akad. Nauk.* 231, 333 (1976)(Figure 1, p.333)

In about 1980, N. Setaka, Y. Sato, and S. Matsumoto et al. of the National Institute for Research in Inorganic Materials in Tokyo, Japan confirmed the activated low-pressure diamond growth from the vapor phase, as shown in Fig. 6.3 and Fig. 6.4. They published papers continuously with their new progress and improvement in this project [17~22], but very few scientists in the West paid attention to their work.

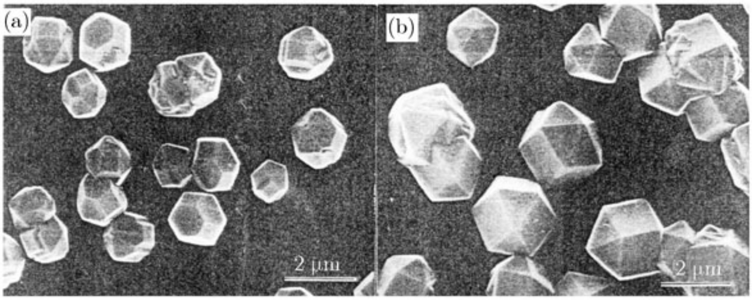


Fig. 6.3 Diamond crystals grown from the vapor phase at low pressure were prepared by Matsumoto, Sato, Setaka et al. with the hot filament method in 1982. Reprinted from S. Matsumoto, Y. Sato, M. Tsutsumi and N. Setaka (Figure 7, p. 3111)[17], (Chapman & Hall Ltd.)

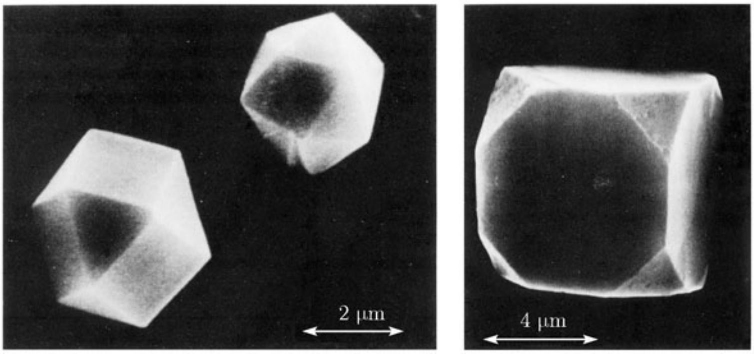


Fig. 6.4 Diamond crystals grown from the vapor phase at low pressure by the hot filament method were published by Setaka in 1989. Reprinted from N. Setaka (Figures 5 and 15a, p. 666 and p. 670)[22] by permission of the Materials Research Society

In 1984, Rustum Roy of Penn State University, USA visited the National Institute for Research in Inorganic Materials in Tokyo. Setaka told him about the authenticity of activated low-pressure diamond growth from the vapor phase initiated in the former Soviet Union, and showed him very beautiful diamond crystals grown at low pressure from the vapor phase by microscopy. The visual evidence was both convincing and stimulating, so on his return Roy started a project on activated CVD diamond with funding from the Office of Naval Research. After a short time he succeeded in producing diamond films. Roy announced his results at a news conference in February 1986. Since then the activated low-pressure diamond synthesis, or activated CVD diamond growth, has become a major focus of research throughout the world. In 1987, K.E. Spear of Penn State University published a review of the history of the activated low pressure diamond process. He said that[23]

> The new diamond technology has a curious history that is a story of scientific caution and communication difficulties. It was perhaps influenced by the mystique that surrounds diamonds — a mystique that has led some scientists to joke of “alchemy” when discussing their (note: former Soviet scientists) investigations. The activated vapor process was first recognized in the Soviet Union, then in Japan, and finally, in the past few years, in the West.

At the beginning of the 1980s, N. Setaka et al., during their confirmation of the achievement of the former Soviet Union scientists, also made improvements to the experimental equipment, process design and so on. For instance, the setup of the hot filament CVD process (HFCVD), as shown in Fig. 6.5, and the setup of the microwave plasma CVD process (MPCVD), as shown in Fig. 6.6, are simple and work stably.

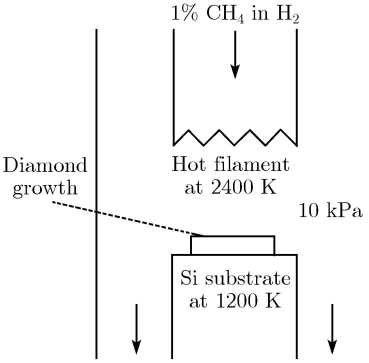


Fig. 6.5 Hot filament process

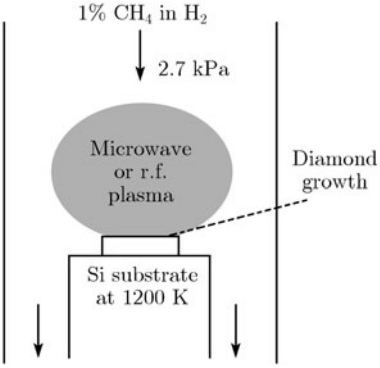


Fig. 6.6 Microwave or r.f. plasma process

The typical diamond growth conditions of the hot filament process are: 1% CH_4 in hydrogen gas mixture is used as carbon sources; 2400 K is for the activation temperature of the hot filament; the refractory metal of molybdenum, tungsten or tantalum is necessary for material of the hot filament; 1200 K is for the Si or Mo substrate temperature; the distance between the hot filament and the substrate is about 10 mm; and 10 kPa (0.1 atmosphere) of the total pressure. The main effect of the hot filament was to produce more hydrogen atoms. The concentration of atomic hydrogen produced is much higher than the equilibrium concentration at the substrate temperature (1200 K), so it is called superequilibrium atomic hydrogen (SAH).

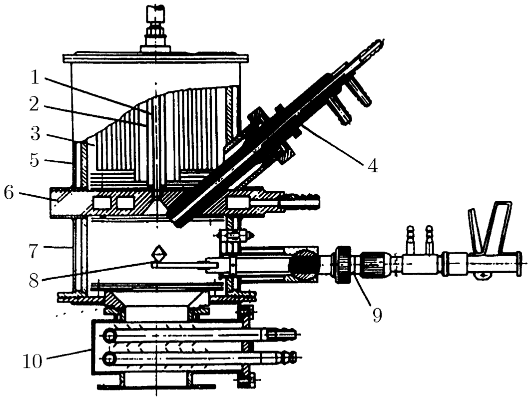


Fig. 6.7 Superequilibrium atomic hydrogen introduced for activated low-pressure diamond growth in the early time at the Institute of Physical Chemistry in the former Soviet Union. (1) tungsten tube for hydrogen supply, (2) tungsten heater, (3) Mo screens, (4) methane supply, (5) bell-jar, (6) water-cooled copper diaphragm, (7) reaction chamber, (8) diamond seed crystal, (9) positioner, (10) water-cooled exhaust. Reprinted from B.V. Spitsyn [14] by permission of The Electrochemical Soc., Inc.

The existence of SAH plays a key role in the new activated low-pressure diamond growth. Figure 6.7 illustrates the apparatus used by B.V. Deryagin, B.V. Spitsyn, D.V. Fedoseev et al. in the early 1970s. Hydrogen and methane were separately introduced into the system. Superequilibrium atomic hydrogen was produced by the tungsten heater at the very high temperature. The importance of superequilibrium atomic hydrogen was very clearly shown.

Another kind of the activated low-pressure diamond syntheses was the process of high gradient chemical transformation reaction (HGCTR), which was also used by Deryagin et al. in the early 1970s. The setup of this process is illustrated in Fig. 6.8. It can be found that the equipment, as shown in both of Fig. 6.7 and Fig. 6.8, was designed for vapor epitaxial growth of single-crystal diamond. Epitaxial layers of diamond had been obtained at that time, but the layers were very thin and with a lot of defects, as shown in Fig. 6.9. That

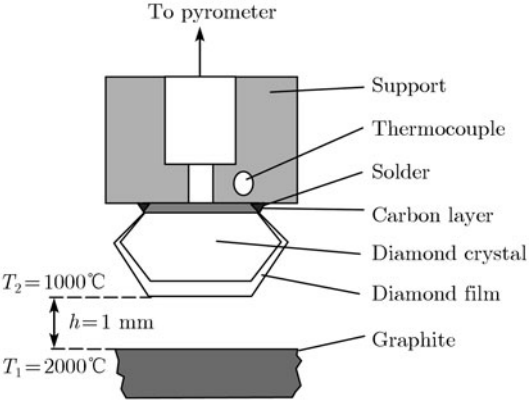


Fig. 6.8 High temperature gradient chemical transport (HGCTR) process (at 1.6 kPa in H_2). Diamond is obtained from graphite. Reprinted from B.V. Spitsyn [14] by permission of The Electrochemical Soc., Inc.

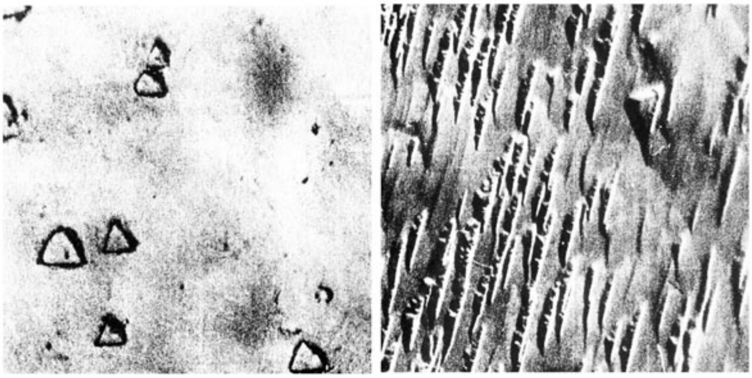


Fig. 6.9 Morphology of epitaxial diamond films on natural diamond substrate, (a) 3.7 μm thick grown on a (111) natural diamond substrate (1200 ×), (b) 2 μm thick grown on a polished natural diamond surface with crystallographic orientation about 5° off the (100) plane (800 ×). Reprinted from B.V. Spitsyn [15] by permission of Elsevier Science. A similar to (b) photograph of epitaxial diamond film on natural diamond substrate had been published in 1971 [11]

is, no big enough gem-quality single-crystal diamonds had been prepared by the activated low-pressure diamond syntheses at that time and even later in the whole 20th century.

There were also many other processes to realize activated low-pressure diamond growth from the vapor phase. Such as plasma jet process, all of them can be used for activation, as shown in Fig. 6.10. The acetylene-oxygen combustion process, as shown in Fig. 6.11, also reached success, and many hydrocarbons, alcohol, methanol, acetone etc. oxygen-containing, and halogen-containing compounds can also be used as the combustion gas instead of acetylene. In fact, any kind of external energy sources and any carbon-containing compounds can be used for the activated low-pressure diamond growth.

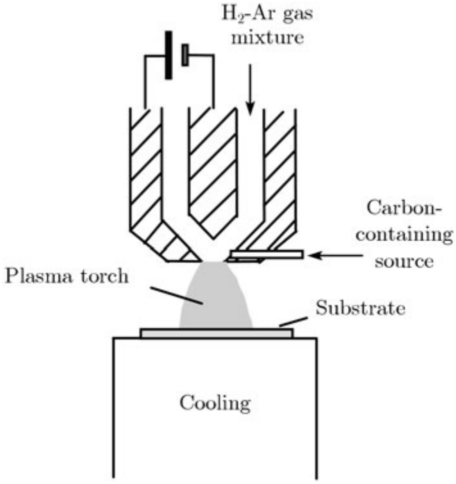


Fig. 6.10 Plasma torch process

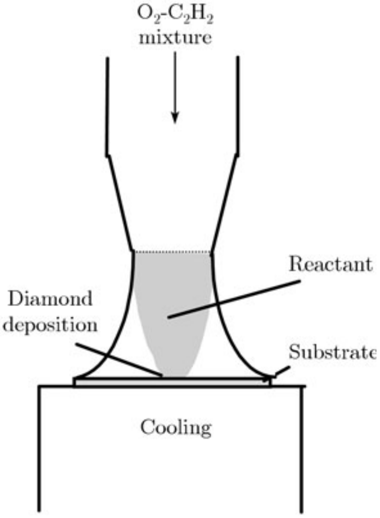


Fig. 6.11 Combustion process

Figure 6.12 shows a closed system in which the transformation from graphite to diamond can be realized. The phenomenon of low-pressure diamond growth and graphite etching taking place simultaneously is difficult for classical thermodynamics to explain. At the beginning inside the closed system there was only hydrogen gas. The plasma was generated by radio-frequency or microwave. During the graphite etching process, molecules of hydrocarbons, methane, ethylene etc. were produced in the gas phase. Hydrocarbon gas molecules might be pyrolyzed under the existence of superequilibrium atomic hydrogen to grow diamond on the nondiamond substrate (such as silicon wafer) or on diamond substrate. Therefore, such a setup can be used to see the typical phenomenon of simultaneous diamond growth and graphite etching.

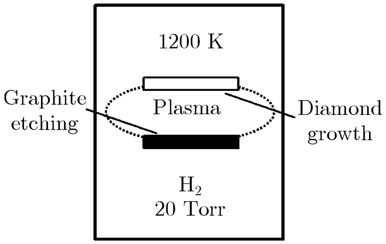


Fig. 6.12 Closed microwave plasma process

In fact, the phenomenon that the diamond growth takes place with simultaneous graphite etching during the existence of superequilibrium atomic hydrogen can be confirmed in any setup of these new diamond processes. If a few graphite particles were put on the surface of diamond growth, under the existence of superequilibrium atomic hydrogen the phenomenon of the low-pressure diamond growth together with the simultaneous graphite etching can be observed. It is well-known that in equilibrium phase diagram of carbon graphite is a stable phase, while diamond is a metastable phase. Therefore some scientists have regarded the new diamond technology as a "thermodynamic paradox", perhaps "violating the second law of thermodynamics".

6.3 Preferential Etching Kinetic Model of SAH

For the phenomenon of the low-pressure diamond growth together with the simultaneous graphite etching during the existence of superequilibrium atomic hydrogen,[14, 19, 24] Spitsyn et al. proposed rather early a kinetic model on the basis of the preferential etching of graphite by superequilibrium atomic hydrogen.[15] The key viewpoints of this model are as follows.

1. If there is only methane in the reactant gas, the methane pyrolysis rates for producing graphite or diamond can be adjusted by selecting the kinetic conditions so that the deposition or growth rate of diamond may be a little higher than the deposition rate of graphite, as shown in Fig. 6.13 (a).
2. If there is only atomic hydrogen, the etching rate of graphite by atomic hydrogen is much higher than the etching rate of diamond, as shown in Fig. 6.13 (b).
3. If both methane and hydrogen with a superequilibrium concentration of atomic hydrogen are used, then due to the combination of both the methane pyrolysis effect and the atomic hydrogen etching effect the situation that the diamond growth rate is positive, while the graphite growth rate is negative (i.e. etching), may be presented as shown in Fig. 6.13 (c).

Figure 6.13 illustrates the key points of the model. Here, *r* and *e* represent the growth rate and the etching rate, respectively. The subscripts "d" and "g" represent diamond and graphite, respectively.

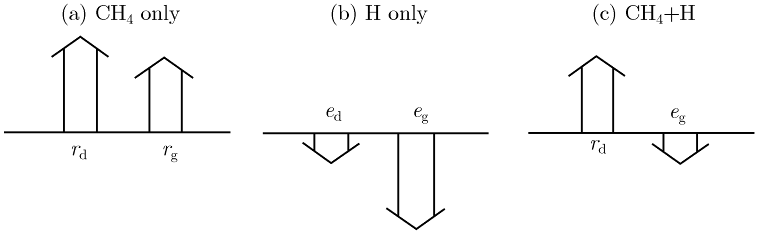


Fig. 6.13 An illustration of preferential etching kinetic model of superequilibrium atomic hydrogen

At that time there was no suitable thermodynamic model, so such a kinetic model could provide some theoretical support or explanation for the new low-pressure diamond process. The shortcoming of that model was that only qualitative explanations could be provided, and it was difficult to make further quantitative calculations.

Note that, at that time there was another kinetic control model, in which the so-called "metastable growth" of diamond is emphasized.[25] The model suggested that during low-pressure diamond growth from the vapor phase the process can be controlled by kinetic effects but not by thermodynamic effects. The key points of view of that kinetic control model may be described as follows.

1. Low-pressure diamond growth from the vapor phase is controlled by kinetic effects but not by thermodynamic effects.
2. Hydrocarbons are of higher chemical potentials.
3. During the pyrolysis of hydrocarbons there is both a possibility to produce graphite and a possibility to produce diamond.
4. At low pressures diamond is a metastable phase and graphite is a stable phase. However, the free energy difference between graphite and diamond is very small, and the barrier between graphite and diamond is high enough. Therefore, if the kinetic parameters are well selected, it will be possible to get diamond but not graphite.

The kinetic control model can be illustrated as shown in Fig. 6.14. From viewpoints 2, 3 and 4, the details of the changes are as follows. Carbon atoms in hydrocarbons are of higher chemical potential energy. During pyrolysis of gaseous hydrocarbons, carbon atoms can drop into both the graphite and diamond phases of lower chemical potential energies. While the energy level of the graphite phase is a little lower, it is still possible to make carbon atoms drop mainly into the diamond metastable phase and to get diamond growth by selecting the kinetic parameters. However, that model could not provide an explanation for diamond growth with the simultaneous graphite etching in the activated low-pressure CVD diamond process. The main problem of the kinetic control model was that kinetics was considered to contradict thermodynamics for activated low-pressure diamond growth from the vapor phase.

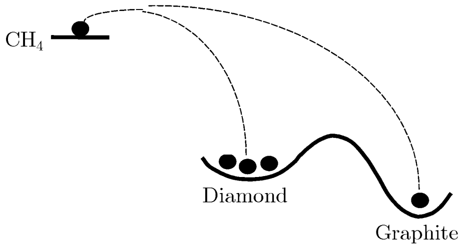


Fig. 6.14 An illustration of kinetic controlled model

It is well known that kinetics from the microscopic viewpoint and thermodynamics from the macroscopic viewpoint always compensate each other. There are only the following possible situations in their relationships.

1. If a change has already kinetically happened in practice, it must also be thermodynamically possible.
2. If a change has been correctly predicted to be thermodynamically impossible, then it should never take place kinetically.

3. If a change is thermodynamically possible, it may be impossible to carry out in practice due to its kinetic rate being too slow, such as there is no reaction to take place at room temperature for the mixture of (2 H_2+O_2) at room temperature.
4. If both of two different reactions are thermodynamically possible, then kinetics can play a deciding role. The kinetic control model must be incorrect because the low-pressure diamond synthesis was in general regarded as thermodynamically impossible. In the case of activated low-pressure diamond growth from the vapor phase which has already taken place in practice, it will not be thermodynamically impossible, or out of thermodynamic control.

In our previous paper (1994) it was pointed out[26]

> There had been an incorrect point of view that the growth of activated CVD diamond is controlled only by kinetics, not thermodynamics. It is incorrect, because kinetics and thermodynamics are compensated each others from different sides, but not against each other. As a basic point of view in physics and chemistry should be that any practical kinetically possible process must be thermodynamically possible without any exception, but, inversely, thermodynamically possible process may be kinetically impossible, or its kinetic rate approximately equals to zero. A famous example is (2 H_2 + O_2) mixture. The formation of water from the mixture is absolutely thermodynamically possible even at room temperature, but in fact the kinetic rate of the water formation reaction nearly equals zero, i.e. the reaction is kinetically impossible at room temperature without a catalyst (such as platinum black) or an initiation (such as a spark).

6.4 Some Thermodynamic Models of the 1980s

In the middle of the 1980s, the activated CVD diamond growth had become a major focus of research throughout the world. A lot of researchers came from different specialties, such as physics, chemistry, mechanics, electronics, materials, and so on. However, the basic thermodynamic question of why diamond can grow under activated low pressure conditions had not yet been satisfactorily answered. In fact, that also presented a challenge to classical thermodynamics.

In the 1980s, the equilibrium phase diagram of carbon was already known, as shown in Fig. 6.15. There was a key puzzle for theoretical researchers. It was well confirmed in classical equilibrium thermodynamics that at low pressures diamond is a metastable phase, while graphite is a stable phase. Why in the new diamond process can "metastable" diamond growth and "stable" graphite etching take place simultaneously at low pressures? Which is more stable, diamond or graphite at such low pressures? In particular, synthetic diamond crystals can be obtained from graphite as the unique carbon source inside a closed system, as shown in Fig. 6.12. That had puzzled more scientists for a long time.

To sum up, during most of the 1980s there was little research from the theoretical viewpoint on this thermodynamic puzzle. In 1989 three theoretical models were proposed. Some scientists had already realized that the activated low-pressure diamond growth system was a nonequilibrium system. Thus different kinds of revisions or assumptions had been made to classical equilibrium thermodynamics, and that resulted in different explanations. However, most scientists are used to using equilibrium thermodynamics, so to cast off such long-term dominant influences of classical equilibrium thermodynamics should spend some time.

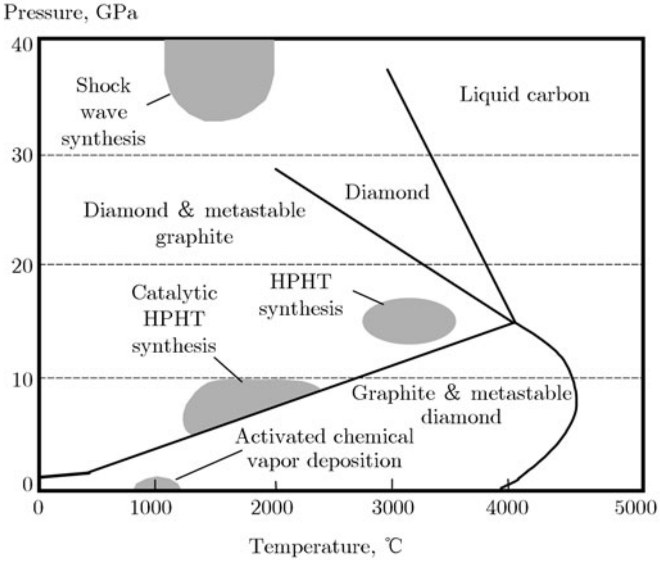


Fig. 6.15 Equilibrium phase diagram of carbon in the late 1980s

6.4.1 *Quasiequilibrium Model*

M. Sommer et al. proposed a quasiequilibrium model in 1989.[27] There is a quasiequilibrium method, which had been proposed and used in research work on oxides.[28] This method includes many kinetic steps and introduced many kinetic parameters. According to the quasiequilibrium method and kinetic theory, conclusions similar to that from equilibrium thermodynamics may be derived. A quasiequilibrium phase diagram of the carbon-hydrogen binary system was obtained, as shown in Fig. 6.16. In fact, this “quasiequilibrium” phase diagram was basically the same as the equilibrium phase diagram of the carbon-hydrogen binary system without substantial difference.

In the second step, Sommer et al. assumed that the desorption coefficient of hydrogen molecules, which were absorbed and underwent a reaction equilibrium on the graphite surface, did not equal 100 % as used in Fig. 6.16, but equals 20%, while the desorption coefficient of hydrogen molecules on the diamond surface equals to 100% and remained unchanged both in Fig. 6.16 and in Fig. 6.17 for activated low-pressure diamond growth from the vapor phase. Based on such an assumption of modification, a modified quasiequilibrium phase diagram, as shown in Fig. 6.17, was obtained. In Fig. 6.17 the graphite phase line shifted to the right side of the diamond phase line, and thus a single diamond solid phase region emerged between the gas-solid phase line of diamond and the gas-solid phase line of graphite. Experimental data for synthetic diamond grown in the hot filament process and reported by Mastumoto were just dropped inside the diamond phase region of the phase diagram.[17]

The success of that model was a simple, clear and even quantitative way to explain experimental phenomena by “modified” quasiequilibrium phase diagram. However, the shortcomings were as follows.

1. The "modified quasiequilibrium phase diagram" was based on an arbitrary assumption.
2. The assumption was not directly connected with the important role of superequilibrium atomic hydrogen.

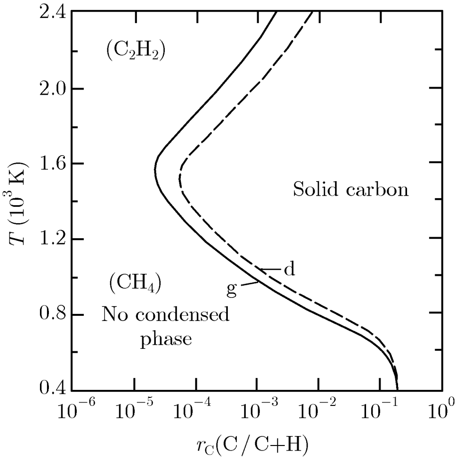


Fig. 6.16 Phase diagrams of the carbon–hydrogen binary system calculated by the quasiequilibrium method at a total pressure of 4.8 kPa. Assuming that the H_2 desorption coefficients on both graphite and diamond surfaces equal 100%. Reprinted from M. Sommer, K. Mui and F.W. Smith (Figure 2, p. 776) [27] by permission of Elsevier Science

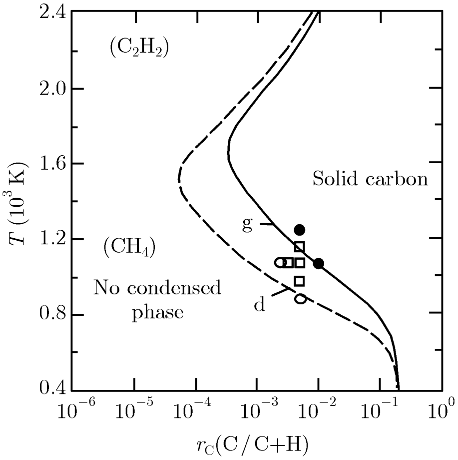


Fig. 6.17 Phase diagrams of the carbon–hydrogen binary system calculated by the revised quasiequilibrium method at a total pressure of 4.8 kPa. Assuming that the H_2 desorption coefficient on graphite is 20% but on diamond is still equal to 100%. Reprinted from M. Sommer, K. Mui and F.W. Smith (Figure 3, p. 777) [27] by permission of Elsevier Science

3. If the substrate temperature were equal to the hot filament temperature in the hot filament process the graphite phase line of the original quasiequilibrium phase diagram

should meet the graphite line of the modified quasiequilibrium phase diagram at the same point; this did not happen. Due to the rigour of thermodynamics, more serious problems could be deduced from these insufficiencies.

6.4.2 Surface Reaction Thermodynamics Model

In 1989, Yarbrough proposed a surface reaction thermodynamics model for the activated low-pressure diamond growth process.[29] The key points of view were as follows.

1. Activated low-pressure diamond growth from the vapor phase was a surface reaction process, so the nature of the reaction might be mainly dependent on the thermodynamic change of the surface reaction.
2. The expression of the basic surface reaction was

$$C_2\ [\text{graphite},\ (0001)] + H^{o} = C_2H\ [\text{diamond},\ (111)]. \qquad (6.10)$$

3. The structures of the graphite (0001) surface and the diamond (111) surface in hydrogen are shown in Fig. 6.18. ΔH_f represented the heat of formation, and H° represented atomic hydrogen.

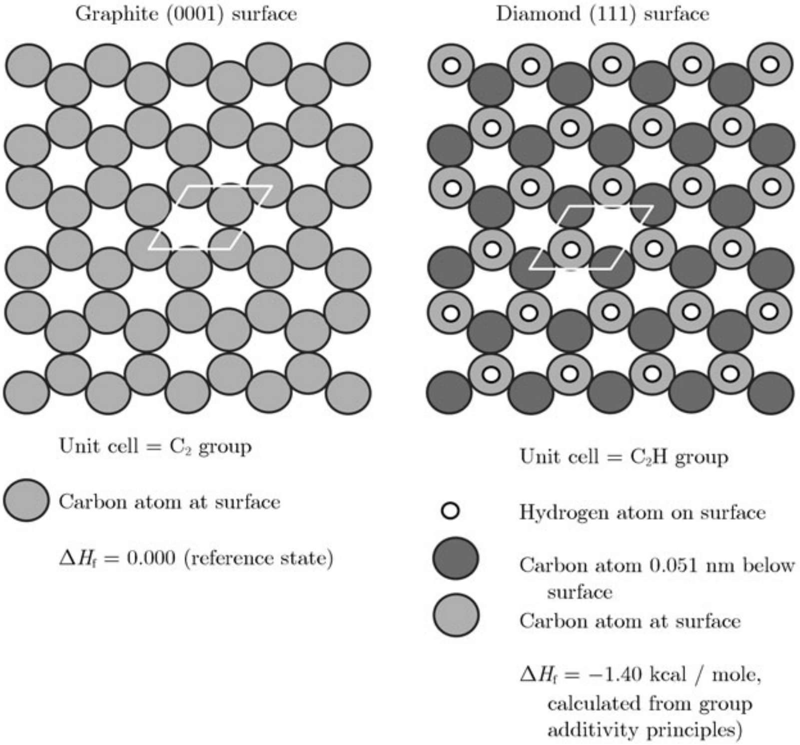


Fig. 6.18 Graphite (0001) surface and diamond (111) surface in hydrogen. Reprinted from W.A. Yarbrough[29] by permission of TMS (The Minerals, Metals & Materials Society)

4. Due to hydrogenation of the diamond surface, the hydrogenated diamond surface C_2H [diamond, (111)] were more stable than the graphite surface C_2 [graphite, (0001)], so the above reaction should go to the right-hand side.

According to that theoretical model Yarbrough estimated the stabilities of the hydrogenated diamond surface and the graphite surface. The conclusion was that the relative stabilities of the hydrogenated diamond surface and graphite surface changed with the concentration of hydrogen atoms, as shown in Fig. 6.19.

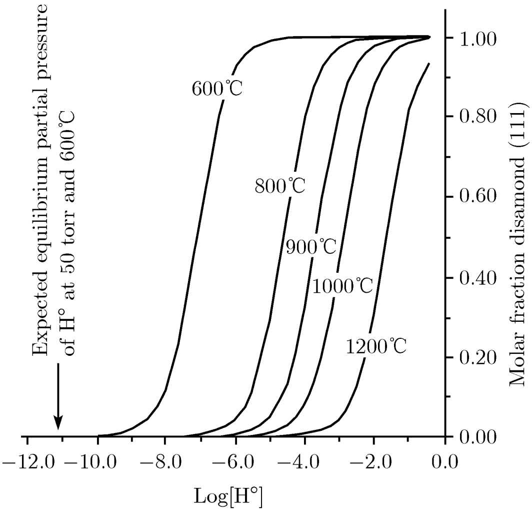


Fig. 6.19 Variation of mole fraction diamond (111) versus atomic hydrogen partial pressure for five different surface temperatures. Reprinted from W.A. Yarbrough [29] by permission of TMS (The Minerals, Metals & Materials Society)

The figure showed that when the concentration of hydrogen atoms was high enough, the diamond phase would be more stable than the graphite phase, while the graphite phase is more stable than the diamond phase at ordinary low-pressure conditions. In the note of Fig. 6.19, Yarbrough wrote:

> Also shown is the expected (equilibrium) partial pressure of atomic hydrogen at the lowest substrate temperature (600°C) for which calculations were made. For comparison the expected partial pressure of atomic hydrogen at 10 Torr and 1000°C is about 1.66×10^{-7} Atm or $\log[H^o] = -6.78$. At 2000 °C and 10 Torr, typical operating conditions in the hot wire synthesis of diamond, the expected partial pressure of atomic hydrogen (at the filament) is 2.09×10^{-3} Atm or $\log\,[H^o] = -2.68$. Clearly if a significant fraction of the atomic hydrogen formed in a plasma or at a hot wire persists to form a steady state partial pressure significantly above its equilibrium value for a given surface temperature, diamond not graphite is the indicated product. At very high surface temperature, extremely high partial pressure of atomic hydrogen would be required for steady state diamond growth.

The advantage of that model was the introduction of the well-known important role of superequilibrium atomic hydrogen into its theoretical treatment. However, the basic shortcoming of that model was that thermodynamics does not allow to arbitrarily change the initial state (graphite) and the final state (diamond) of the research system. However, in the surface reaction model the initial state had been changed into graphite surface, and the final state had been changed into diamond surface. Any surface does not exist without its substrate. If the surface together with its substrate is drawn out, then it will be found that the surface reaction thermodynamics model is not reasonable.

"Surface reaction (a)": If reactant of "graphite (0001) surface on graphite (0001) substrate" reacts with atomic hydrogen by "surface reaction (a)", then the product of so-called "diamond (111) surface on graphite (0001) substrate" would be obtained. That is, the reactant is real graphite with (0001) orientation, but the "product" does never exist, as shown in Fig. 6.20.

$$\frac{\mathrm{C_2[Gra(001)_{Sur}]}}{\text{on }\mathrm{Gra(001)_{Sub}}} + \mathrm{H}^\circ = \frac{\mathrm{C_2H[H{-}GIa(001)_{Sur}]}}{\text{on }\mathrm{Gra(001)_{Sub}}}$$

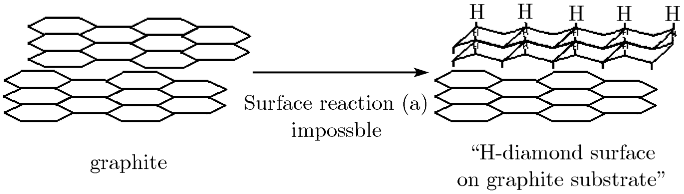


Fig. 6.20 Surface reaction (a) – impossible to take place

"Surface reaction (b)": If reactant of so-called "graphite (0001) surface on diamond (111) substrate" could react with atomic hydrogen by "surface reaction (b)", then the product of "hydrogenated diamond (111) surface on diamond (111) substrate" would be obtained. That is, the product would be real hydrogenated diamond with (111) orientation, but the "reactant" does never exist, as shown in Fig. 6.21.

$$\frac{\mathrm{C_2[Gra(001)_{Sur}]}}{\text{on }\mathrm{Gra(111)_{Sub}}} + \mathrm{H}^\circ = \frac{\mathrm{C_2H[H{-}GIa(001)_{Sur}]}}{\text{on }\mathrm{Gra(111)_{Sub}}}$$

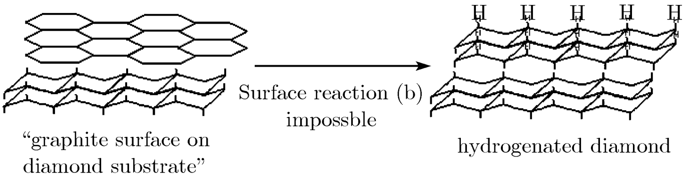


Fig. 6.21 Surface reaction (b) – impossible to take place

"Case (c)": If reactant of "graphite (0001) surface on graphite (0001) substrate" could react with atomic hydrogen as "case (c)", and the product were "hydrogenated diamond (111) surface on diamond (111) substrate", then the "case (c)" will not be surface reaction but bulk reaction instead. Because only one surface could react with atomic hydrogen with a little chemical energy, the whole bulk will not be able to transform, as shown in Fig. 22.

Besides, surface thermodynamic data were estimated by Yarbrough himself. In thermodynamics calculation, thermodynamic data should not be estimated or changed arbitrarily. Revised thermodynamic data might get arbitrary results. Yarbrough had also mentioned the problem.

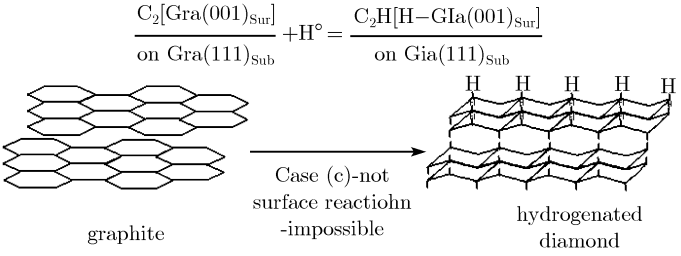


Fig. 6.22 Case (c) – not surface reaction, impossible to take place

6.4.3 Defect-Induced Stabilization Model

In 1989, Y. Bar-Yam et al. also proposed a defect-induced stabilization model for the activated low-pressure diamond film deposition in *Nature*.[30] Bar-Yam et al. said:

> Under such growth conditions diamond should be unstable relative to graphite, yet diamond is formed in practice.

The basic points of view of their model were as follows.

1. During the activated low-pressure diamond growth, high vacancy concentrations were included in the prepared diamonds. Therefore, a comparison of stabilities between diamond with high vacancy concentrations and graphite with high vacancy concentrations should be made.
2. It was assumed that the formation free energy for diamond and graphite with a concentration n of vacancies was $F = F_b + nF_d$. Here, F_b is the binding energy of the materials with negligible defect concentrations and F_d is the free energy of formation of a defect or vacancy. Preferential diamond growth would occur if the formation energy of vacancies was smaller in diamond and

$$n > \frac{F_b(\text{dia}) - F_b(\text{gra})}{rF_d(\text{gra}) - F_d(\text{dia})} \tag{6.11}$$

where r is the ratio of vacancies in graphite and diamond. Thus, a significant energy difference in the energies of defect formation might change the balance between the close binding energies of graphite and diamond.

The vacancy formation energy in diamond and in graphite was calculated by them as 7.2 eV and 7.6 eV, respectively. It is known that the difference in the cohesive energies of diamond and graphite is about 0.03 eV. According to Eq. (6.11), assuming $r \approx 1.8\%$, vacancies at the growth surface would stabilize diamond with respect to graphite, so in the activated low pressure CVD process it was favorable to grow diamond with a lot of defects instead of growing graphite with a lot of defects.

Based on these points of view the energy levels may be illustrated in Fig. 6.23. Obviously, that theoretical model could not explain why perfect graphite crystals were not grown with the lowest energy level. In fact, as shown in Fig. 6.2, Fig. 6.3 and Fig. 6.4, nice crystals can also be obtained by the activated low-pressure CVD diamond process, and even diamond epitaxial layers on a diamond substrate can also be prepared. All of these were difficult to explain by the defect-induced stabilization model.

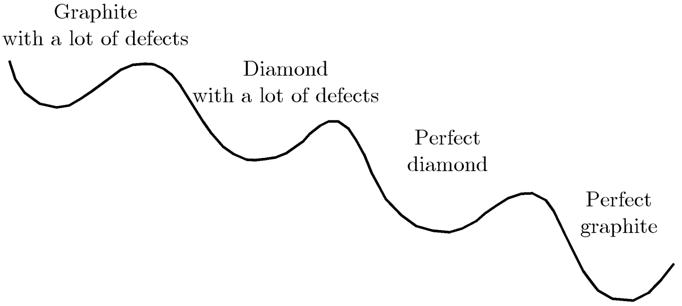


Fig. 6.23 An illustration of the changes of energy level introduced by defects

In the February 1990 issue of the journal *Science*, Yarbrough and Messier published their paper "*Current Issue and Problems in the Chemical Vapor Deposition of Diamond*". After a brief discussion on thermodynamic models already published at that time (i.e. quasiequilibrium model, surface reaction thermodynamics model, defect-induced stabilization model etc.) they made the comment that[31]

> All of these formulations suffer from being inconsistent with at least some experimental results.

That was an objective comment on these models, including the model proposed by Yarbrough himself, and reflected the theoretical situation in activated CVD diamond research. In fact, the basic problem was due to the activated nonequilibrium conditions for the activated low-pressure diamond growth being out of the applicability framework of classical thermodynamics.

6.5 Thermodynamic Coupling Model

In April 1990, the author, as the first speaker at *the 8th International Conference on Thin Films* (San Diego, US), together with Jan-Otto Carlsson proposed a thermochemical model or called chemical pump model.[32] The key points of view in this model were that a large amount of energy is released by the association of superequilibrium atomic hydrogen to produced molecular hydrogen; carbon atoms of graphite phase are transferred to the diamond phase of a higher potential energy by addition reaction between atomic hydrogen and graphite; and the process is similar to a "chemical pump", increasing chemical potential energy of graphite, and promoting the transfer from graphite to diamond. A little later, the author found that similar reactions in the living body had already been called "thermodynamic coupling". Showing respect to the term used by former scientists, after 1994 our "chemical pump model" was renamed "thermodynamic coupling model", and the model was continuously improved.[26,32~39] The term "pump", such as "sodium, potassium ion pump" etc., is still used by many biochemists and so the term "chemical pump" can still be used. However, from a thermodynamics viewpoint, the term thermodynamic coupling is better.

Thermodynamic coupling model can be simply expressed by the following set of reactions, i.e. the activated low-pressure diamond growth from the vapor phase at least includes the following two main reactions:

$$\begin{gathered}\mathrm{C(gra)} = \mathrm{C(dia)},\\ (T, p \leqslant 10^5\ \mathrm{Pa})\\ (\Delta G_1)_{T,p} > 0, \Delta_\mathrm{i} S_1 < 0.\end{gathered} \tag{6.12}$$

$$\begin{gathered}\mathrm{H}* = 0.5\mathrm{H}_2,\\ (T_\mathrm{activation} \gg T, p \leqslant 10^5\ \mathrm{Pa})\\ (\Delta G_2)_{T,p} \ll 0, \Delta_\mathrm{i} S_2 \gg 0.\end{gathered} \tag{6.13}$$

The change of the whole system, or the total reaction should be the sum of Eq. (6.12) and Eq. (6.13). At the same time, the ratio of reaction Eq. (6.13) to reaction Eq. (6.12) $\chi = r_2/r_1$ should be under consideration. If χ is not too small, then the total reaction can be obtained.

$$\begin{gathered}\mathrm{C(gra)} + \chi\ \mathrm{H}* = 0.5\chi\ \mathrm{H}_2 + \mathrm{C(dia)},\\ (T_\mathrm{activation} \gg T, p \leqslant 10^5\ \mathrm{Pa})\\ (\Delta G)_{T,p} \leqslant 0, \Delta_\mathrm{i} S \geqslant 0.\end{gathered} \tag{6.14}$$

Here, C (gra), C (dia) and H* represent graphite, diamond, and superequilibrium atomic hydrogen, respectively. $(\Delta G_1)_{T,p}$, $(\Delta G_2)_{T,p}$ and $(\Delta G)_{T,p}$ represent molar Gibbs free energy changes of Eq. (6.12), Eq. (6.13), and the total reaction Eq. (6.14), respectively. $\Delta_\mathrm{i} S_1$, $\Delta_\mathrm{i} S_2$ and $\Delta_\mathrm{i} S$ represent molar entropy productions of Eq. (6.12), Eq. (6.13), and the total reaction Eq. (6.14), respectively. The condition of "χ is not too small" means the concentration of superequilibrium atomic hydrogen must be over a certain value, or say, the activated temperature (i.e. hot filament temperature) must be much bigger then that of substrate temperature ($T_\mathrm{activation} >> T$). Every expert on the activated low-pressure diamond growth knows that in hot filament process, the temperature of hot filament should not be too low. If the substrate temperature is 1000–1200 K, the alloy of platinum (m.p. 1600°C, pure platinum: m.p. 1772°C) should not be used as filament. Otherwise, the low-pressure diamond growth will not be stable. Therefore, the filament temperature is usually controlled at the higher than 2000 K temperature, and tungsten (m.p. 3410°C) or tantalum (m.p. 2996°C) of high melting point is used as hot filament material. The expression of $(\Delta G)_{T,p} \leqslant 0$ shows clearly that the total reaction Eq. (6.14) should go to the right side, i.e. if the concentration of superequilibrium atomic hydrogen is high enough, then the reaction between graphite and superequilibrium atomic hydrogen should produce molecular hydrogen and diamond. Therefore, the new activated low-pressure diamond growth process completely agrees with the fundamental laws of thermodynamics. There is nothing "violating the second law of thermodynamics.

For convenience of quantitative discussion and understanding, under the typical condition of hot filament process: p = 10 kPa (0.1 Atm) substrate temperature T = 1200 K, in 1 % CH_4 in hydrogen mixture, and hot filament: $T_\mathrm{activation}$ = 2400 K, Eq. (6.12), Eq. (6.13) and Eq. (6.14) can be rewritten as follows.

$$\begin{gathered}\mathrm{C(gra)} = \mathrm{C(dia)}, \qquad (6.15)\\ (T = 1200\ \mathrm{K}, p = 10\ \mathrm{kPa})\\ (\Delta G_1)_{T,p} = 6.96\ \mathrm{kJ}, \Delta_\mathrm{i} S_1 = -5.80\ \mathrm{J\ K^{-1}}.\\ \mathrm{H}* = 0.5\mathrm{H_2}, \qquad (6.16)\\ (T_\mathrm{activation} = 2400\ \mathrm{K}, T = 1200\ \mathrm{K}, p = 10\ \mathrm{kPa})\\ (\Delta G_2)_{T,p} = -112.72\ \mathrm{kJ},\ \Delta_\mathrm{i} S_2 = 93.93\ \mathrm{J\ K^{-1}}.\end{gathered}$$

For the whole system or the total reaction,

$$\begin{gathered}\mathrm{C(gra)} + \chi\ \mathrm{H}* = 0.5\chi\ \mathrm{H_2} + \mathrm{C(dia)},\ (\chi = r_2/r_1 = 0.28) \qquad (6.17)\\ (T_\mathrm{activation} = 2400\ \mathrm{K}, T = 1200\ \mathrm{K}, p = 10\ \mathrm{kPa})\\ (\Delta G)_{T,p} = -24.60\ \mathrm{kJ},\ \Delta_\mathrm{i} S = 20.50\ \mathrm{J\ K^{-1}}.\end{gathered}$$

The calculation method and data sources will be discussed in detail in the next chapter.

The above-mentioned reactions, Eq. (6.15), Eq. (6.16) and Eq. (6.17), are in agreement with the phenomenon of thermodynamic coupling discussed in detail in the Chapter 4. $(\Delta G_1)_{T,p} > 0$ or $\Delta_\mathrm{i} S_1 < 0$ shows that the reaction Eq. (6.15) is a nonspontaneous reaction, which is impossible to take place alone; while $(\Delta G_2)_{T,p} \ll 0$ or $\Delta_\mathrm{i} S_2 \gg 0$ shows that reaction Eq. (6.16) is a very strong spontaneous reaction, which can not only take place alone but also promote reaction Eq. (6.15) to the right side together. That is, the reaction between graphite and superequilibrium atomic hydrogen to produce molecular hydrogen and diamond, i.e. reaction Eq. (6.17), agrees with all of the criteria of thermodynamic coupling $[(\Delta G_1)_{T,p} > 0, (\Delta G_2)_{T,p} < 0\ \&\ (\Delta G)_{T,p} \leqslant 0]$ or $[\Delta_\mathrm{i} S_1 < 0, \Delta_\mathrm{i} S_2 > 0\ \&\ \Delta_\mathrm{i} S \geqslant 0]$. Here, $(\Delta G)_{T,p} = [(\Delta G_1)_{T,p} + \chi(\Delta G_2)_{T,p}]$ or $\Delta_\mathrm{i} S = (\Delta_\mathrm{i} S_1 + \chi \Delta_\mathrm{i} S_2)$. Therefore, the reaction between graphite and superequilibrium atomic hydrogen to produce molecular hydrogen and diamond under low pressure completely agrees with the second law of thermodynamics.

In 2001, Gogosti et al. in *Nature* pointed out: [40]

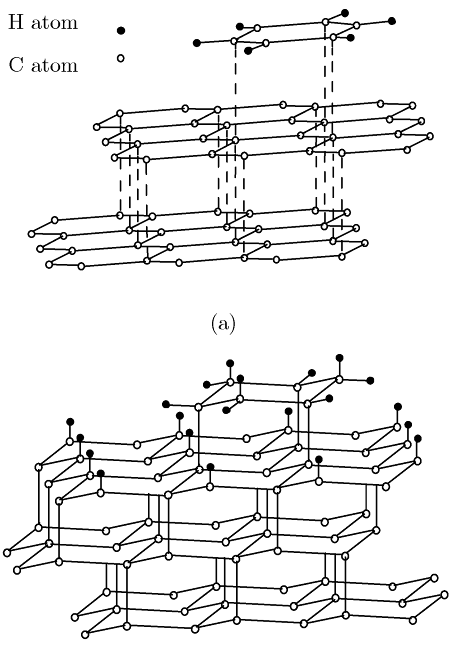
> ···the thermodynamic coupling model that can be used to explain the formation of diamond in reaction (1) and (2) have been well developed.

The cited paper was our 1994 paper published in *J. Electrochemical Society* by the author et al. [26]

6.6 Mechanism of Thermodynamic Coupling in Low-Pressure Diamond Growth

The hybridization states of the carbon atoms in graphite and diamond are different, in $\mathrm{sp^2}$ valent state and $\mathrm{sp^3}$ valent state, respectively. Clusters on the graphite and diamond surfaces are also different during the growth process. The clusters on the (0001) graphite surface in a hydrogen ambiance can be regarded as aromatic polycyclohydrocarbons, whereas the clusters on a (111) diamond surface in the same ambiance are aliphatic ones.

For brevity, only the simplest six-member ring clusters are discussed. The simplest six-member aromatic hydrocarbon is benzene, and cyclohexane is the simplest six-member aliphatic hydrocarbon. A single six-member cluster on a graphite surface in hydrogen can then be treated as a benzene molecule, which is physically adsorbed on the graphite surface, as shown in Fig. 6.24(a). However, a single six-member cluster on a (111) diamond surface in hydrogen can be considered as a cyclohexane molecule, which has adsorbed chemically on the surface, reacted with the diamond hydrogenated surface and released three hydrogen molecules, as shown in Fig. 6.24(b). The graphite clusters are unsaturated. On acceptance of superequilibrium atomic hydrogen (SAH) more saturated hydrocarbons (cyclohexane, cyclohexene, decahydronaphthalene, etc.) are formed. The atomic hydrogen provides sufficient energy for decomposition of the clusters into smaller fragments (CH_4, C_2H_2, C_2H_4, etc.). This results in etching of the graphite clusters as well as in etching of the graphite itself. However, hydrogen shows only weak etching effect or no etching effect on diamond.



(a)

(b)

Fig. 6.24 The cluster on the graphite surface is benzene-like, whereas the cluster on the diamond surface is cyclohexane-like. (a) Cluster on a (0001) graphite surface and (b) cluster on a (111) diamond surface in hydrogen

The favorable influence of atomic hydrogen on diamond growth has been discussed in many papers in the literature. Here, only a few additional comments of the relevance to this study are given. A hydrogenated (111) diamond surface contains a lot of tertiary carbon atoms (the tertiary carbon atom is defined as a carbon atom with three C—C σ single bonds and one C—H σ single bond). On impingement of atomic hydrogen the tertiary carbon atoms become free radicals and molecular hydrogen is formed. The heats of formation of a free tertiary carbon radical, molecular hydrogen and atomic hydrogen are 33 kJ mol^{-1} (based on t-butyl), 0 and 218 kJ mol^{-1} at 298.15 K, respectively. Thus free tertiary carbon radicals are easily formed. These radicals stabilize the (111) surface somewhat (energy decrease). With an increased stability of this surface the formation of gaseous hydrocarbons (etching) is reduced. The tertiary carbon radicals on the surface may act as sinks for hydrocarbon radicals supplied from the vapor and thus support diamond growth.

For a (0001) graphite surface the situation is completely different. There are no tertiary, secondary (the secondary carbon atom is defined as a carbon atom with two C—C σ single bonds and two C—H σ single bonds) or primary carbon atoms (the primary carbon atom is defined as a carbon atom with one C—C σ single bond and three C—H σ single bonds), but a big π bond exists. As discussed above a six-member cluster on a graphite surface in an atomic hydrogen ambiance is benzene-like. The simplest radical of benzene is phenyl (C_6H_5). The heat of formation of a free phenyl radical is 335 kJ mol^{-1} at 298.15 K, which is much higher than the heat of formation of a tertiary carbon radical on a (111) diamond surface. Therefore, the existence of atomic hydrogen does not promote growth on the graphite surface.

In summary, the effect of atomic hydrogen in concentration higher than the equilibrium concentration (superequilibrium atomic hydrogen) can be described in the following way.

1. Graphite surface is easily etched. The superequilibrium atomic hydrogen increases the energy of the surface and transforms it into an unstable state, i.e. a strong chemical pump effect on the graphite surface is obtained. Growth on the graphite surface requires free surface radicals. However, the energy needed to form such radicals, for instance a phenyl radical, is high. This makes growth difficult on graphite.
2. Diamond surface with its tertiary carbon atoms is very different from a graphite surface. Weak etching on the diamond surface in atomic hydrogen occurs. Free surface radicals are more easily formed on diamond than on graphite, promoting growth on diamond. Therefore, the etching and growth promoting effects of superequilibrium atomic hydrogen on the diamond surface nearly offset each other.

The overall effect of superequilibrium atomic hydrogen from an energy point of view is to increase the energy of graphite and to change the energy of diamond slightly. Superequilibrium atomic hydrogen plays a role of changing the relative chemical potential energy and the original stable graphite phase to a metastable phase.

Thermodynamic coupling including chemical reaction(s) is usually called reaction coupling. The substantial effect of thermodynamic coupling is that nonspontaneous reaction(s) or process(es) might be promoted by another energy-release spontaneous reaction(s) or process(es) in the same system at the same time. A lot of examples can be found in life sciences. Due to the complexity, thermodynamic coupling of biological systems could practically be presented by a pair of double arrows. The mechanism of the activated low-pressure diamond growth from the vapor phase can also be presented by a pair of double arrows, as shown in Fig. 6.25(a). Due to simplicity of nonliving system, the mechanism of the activated low-pressure diamond growth from the vapor phase could be described in detail, as shown in Fig. 6.25(b).

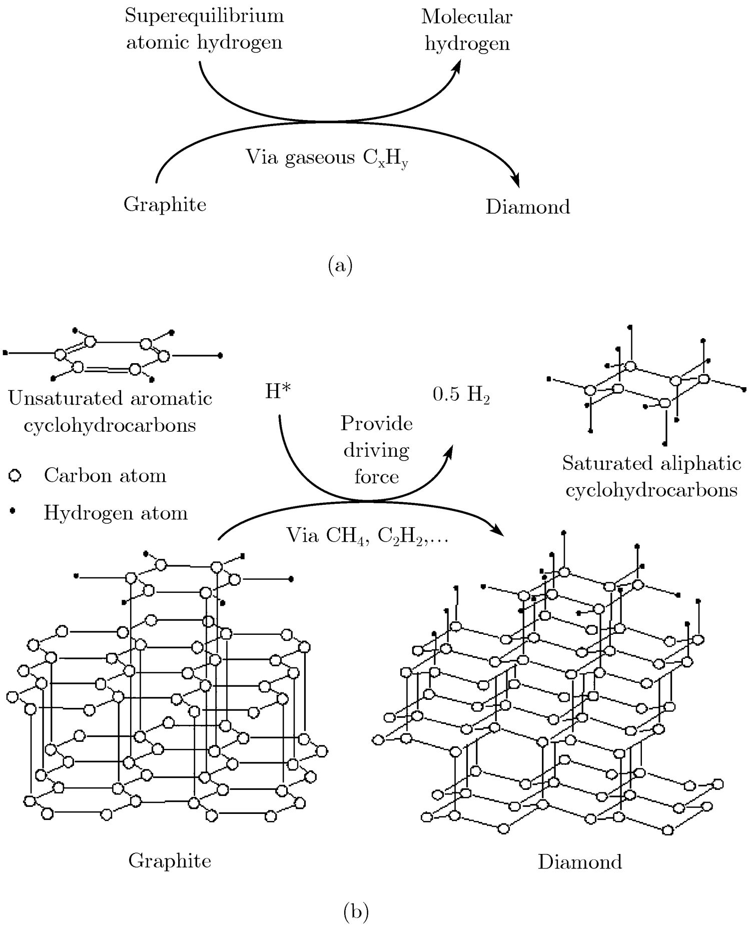


Fig. 6.25 Mechanism for diamond growth at activated low pressure
(a) is a simple expression of (b)

The limitation of thermodynamics is that it only provides information about reaction possibility but not inevitability to realize the reaction. The practical reality should be based on reasonable kinetic mechanism. Therefore, after the discussion above about possibility and rationality of thermodynamic coupling, the detailed kinetic mechanism of thermodynamic coupling should also be discussed. Under the activated low-pressure conditions, from methane etc., hydrocarbons and hydrogen gas mixture diamond, but not graphite, can be grown; or directly from graphite via gas phase, diamond can be grown. All of them are typical organic reactions. Figure 6.25 shows only one of possible idealized mechanisms. In fact, the cluster on graphite surface is usually composed of more carbon atoms (not only 6 carbon atoms). The structure of the cluster on graphite surface is very similar to that of unsaturated aromatic hydrocarbons (e.g. benzene) with sp^2 configuration and big π bonds, as shown at the upper left corner of the figure. The structure of the cluster on diamond surface is very similar to that of saturated aliphatic hydrocarbons (e.g. cyclohexane) with sp^3 configuration and σ bonds, as shown at the upper right corner of the figure. In the gas phase there are mainly methane, acetylene etc. but not benzene or cyclohexane etc. with 6 carbon atoms. Due to the big activity of reaction with unsaturated graphite, superequilibrium atomic hydrogen easily reacts with carbon clusters on graphite surface via addition-pyrolysis reaction to produce corresponding high concentrations of small molecule hydrocarbons (such as methane, acetylene, ethylene, methyl radical, ethyl radical etc.). That is, the gas-solid phase line of the activated graphite shifts toward the right higher carbon concentration side in nonequilibrium phase diagrams, which will be discussed in the next chapters. Concentrations of these gas phase hydrocarbons are much higher than the lowest equilibrium concentrations of hydrocarbons for producing diamond and molecular hydrogen. Therefore, the saturated cluster on diamond surface for diamond to grow up could be formed together with the release of molecular hydrogen. The reversed process is difficult to take place, because carbon atoms on diamond surface are of the saturated sp^3 configuration, which is difficult to react with superequilibrium atomic hydrogen. Many experimental demonstrations in the literature have confirmed that increase of atomic hydrogen concentration has no obvious influence on the etching rate of diamond, while increase of atomic hydrogen concentration shows very obvious influence on the etching rate of graphite. Therefore, during the existence of superequilibrium atomic hydrogen, the formation of graphite nucleus with unsaturated sp^2 configuration is of less possibility. The whole process of coupling is of single direction. After 1995 the discussion more clearly emphasized the key effect of superequilibrium atomic hydrogen, which was presented by H* (with a superscript of *), as shown in Fig. 6.25 (b), instead of H which was used by the author during 1990–1994. If atomic hydrogen is of the equilibrium concentration, then ΔG_2 will be equal to zero, and with no effect from nonequilibrium thermodynamic coupling and driving force to stabilize the low-pressure diamond growth from the vapor phase. The detailed discussion on thermodynamic coupling into the atomic level, as shown in Fig. 6.25 (b), may be unique among all thermodynamic coupling up to now. It is hoped that this will be helpful for understanding reaction coupling and negative entropy production (i.e. nonspontaneous) reactions.

6.7 Other Thermodynamic Models in 1990s for Low-Pressure Diamond Growth

In 1994, a "unified barrier model" was proposed by Y.-F. Zhang. [41] It was pity that in the derivation of that model a "1 + 1 = 1" type mistake had been introduced, meaning that its conclusion was directly against the basic definition of catalyst and the fundamental laws of thermodynamics. Therefore, the author et al. had to publish a paper in the same journal to make a correction. [42]

In 1996, N.M. Hwang et al. proposed a "charged cluster model". [43~44] They believed that

> When the carbon cluster is sufficiently small, the capillary pressure built up inside the cluster can be high enough to make diamond more stable than graphite.

However, that model could not explain the fact that diamond crystals could grow up to micron or millimeter sizes under the activated low-pressure conditions.

In 1998, W. Piekarczyk proposed a "under-saturated crystal growth model". At the beginning he did not agree with our thermodynamic coupling model, but later he emphasized the similarities between the conclusions of both models.[45~46] In fact, however, these two models are different.

6.7.1 Unified Barrier Model

The "unified barrier model" had proposed "a unified barrier", as shown in Fig. 6.26,[41]

> which is thought to be a general mechanism for all these experiments

on synthetic diamond processes, including superhigh-pressure high-temperature processes (HPHT processes), high-pressure catalyst processes (HP catalyst processes), and activated low-pressure CVD processes. The equilibrium point between graphite and diamond was estimated by an approximate calculation of diamond and graphite forward and reverse reaction rates.

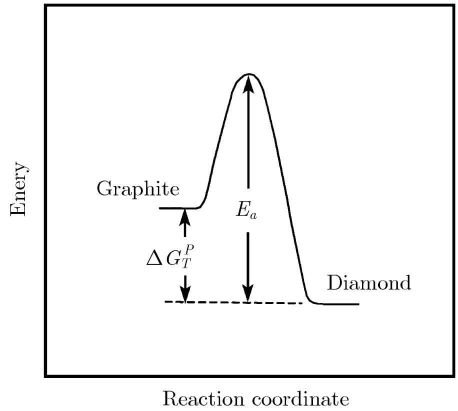


Fig. 6.26 The barrier illustrated in the original paper of the unified barrier model[41]

However, during derivation of forward and reverse reaction barriers a "1 + 1 = 1" type mistake (in fact, a mistake of $\Delta G_T^p + \Delta G_T^p = \Delta G_T^p$) had been introduced. Therefore, that model came to an incorrect result that a catalyst could only accelerate the forward reaction rate but not the reverse reaction rate. In regard to such kinds of conclusion, physical chemistry books had already clearly pointed out a violation against the fundamental laws of thermodynamics. The role of catalyst is to decrease the reaction barrier, so it must accelerate both the forward and the backward reactions. Catalyst does not change the final equilibrium point. Otherwise, the first kind of perpetual machine could be designed. For instance, W. J. Moore in his book "*Physical Chemistry*" (1976) wrote: (Note: The emphasis was made by Moore himself.)(Chap, 9, p.408)[8]

> Wilhelm Ostwald was the first to emphasize that a catalyst influences the rate of a chemical reaction but has no effect on the position of equilibrium. His famous definition was, 'A catalyst is a substance that changes the velocity of a chemical reaction without itself appearing in the end products.' Ostwald's proof was based on the First Law of Thermodynamics. Consider a gas reaction

> that proceeds with a change in volume. The gas is confined in a cylinder fitted with a piston; the catalyst is in a small receptacle within the cylinder, and can be alternately exposed and covered. If the equilibrium position were altered by exposing the catalyst, the pressure would change, the piston would move up and down, and a perpetual-motion machine would be available. Since a catalyst changes rate but not equilibrium, it must accelerate the forward and reverse reactions in the same proportion.

Unified barrier model had also confusedly called completely different processes of the high pressure catalyst process and the activated low-pressure vapor growth process by the common obscure term of "the metal and hydrogen catalyst-assisted diamond synthesis". Therefore, the derivations and results of that model were imbroglio. We have to list the correct barriers in Fig. 6.27, Fig. 6.28 and Fig. 6.29.

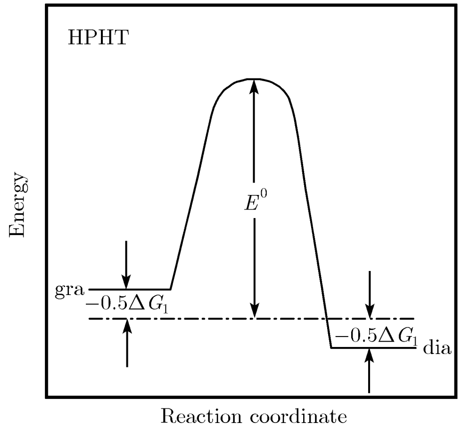


Fig. 6.27 Correct barriers for the super-high-pressure high-temperature synthetic diamond process

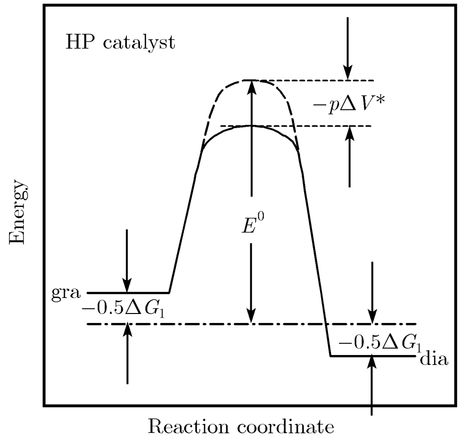


Fig. 6.28 Correct barriers for the high-pressure catalyst synthetic diamond process

If barriers practically used in the unified barrier model are drawn in figures, then unreasonable discontinuous reaction barrier curves are obtained, as shown in Fig. 6.30 and Fig. 6.31. In order to get a continuous reaction barrier curve, if the right part of Fig. 6.30

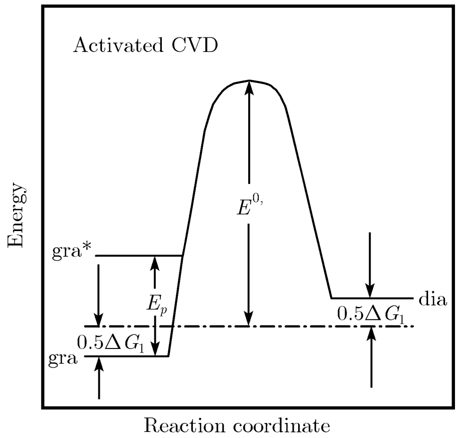


Fig. 6.29 Correct barriers for activated low-pressure CVD synthetic diamond process

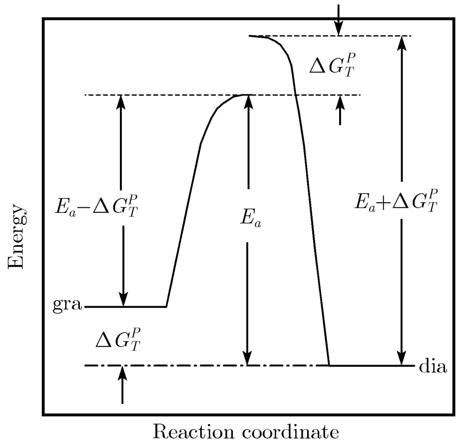


Fig. 6.30 Incorrect barrier used in practical calculations by the unified barrier model for the super-high-pressure high-temperature process

could be lowered by ΔG_T^p, then reaction barrier between graphite and diamond would be 2 ΔG_T^p but not ΔG_T^p as in the original definition of barrier. The high pressure catalyst process and the activated low-pressure CVD process had been confused by the common obscure term of "the metal and hydrogen catalyst-assisted diamond synthesis" in unified barrier model, as shown in Fig. 6.31. Even if the right part of Fig. 6.31 could be lowered by ΔG_T^p, a continuous reaction barrier curve would not be obtained.

For clarification, we had to make a complete correction for derivation, to redraw all of related figures, and to publish our results in the same journal (*J. Materials Research*). The details will not be repeated here. It can be referred in this journal, in the 2000 Chinese monograph "*Stationary nonequilibrium phase diagrams — thermodynamics for low-pressure diamond growth from the vapor phase*", or in the 2002 English monograph "*Nonequilibrium Nondissipative Thermodynamics — With Application to Low-Pressure Diamond Synthesis*", written by the author.[36~38,42]

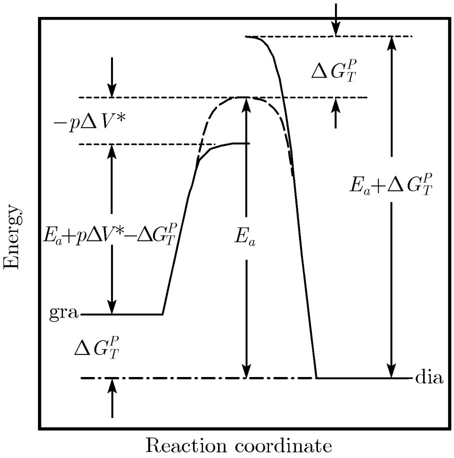


Fig. 6.31 Incorrect barrier used in practical calculations by the unified barrier model for the high-pressure catalyst process and the activated low-pressure CVD process (these two processes are confused in the unified barrier model as "metal and hydrogen catalyst-assisted process")

6.7.2 Charged Cluster Model

In 1996, N. M. Hwang et al. continuously published papers in *J. Crystal Growth*, and proposed a "charged cluster model". [43, 44] The basic premise of that model stemmed from some erroneous thinking about the activated low-pressure diamond growth from the vapor phase. Hwang et al. wrote a series of misguided words. For example, they considered that the well-known experimental observation of

> the preferential etching of graphite over diamond during the existence of superequilibrium atomic hydrogen is a hypothesis···, which is contradictory to the well-established stability of graphite with respect to diamond.

In particular, they misunderstood another well-confirmed experimental observation that

> diamond can be synthesized with graphite as the only source of carbon and diamond is deposited with simultaneous etching of graphite, ···has a critical handicap of contradicting the accepted phase diagram of carbon and thereby violating the second law of thermodynamics.

Moreover, they wrote:

> The experimental observation of diamond deposition with simultaneous graphite etching sounds puzzling in view of the second law of thermodynamics.

The activated low-pressure CVD diamond process even seemed in their papers to be regarded as a "thermodynamic paradox" perhaps "violating the second law of thermodynamics". Correspondingly, Hwang et al. wrote:

> Charged carbon nuclei clusters, which are expected to form in the gas phase, are suggested to be responsible for the formation of diamond. When the carbon cluster is sufficiently small, the capillary pressure built up inside the cluster can be high enough to make diamond more stable than graphite. ··· The gas activation process produces charges such as electrons and ions, which are

> energetically strong heterogeneous nucleation sites for the supersaturated carbon vapor, leading to the formation of charged nuclei clusters ⋯.

and

> Hwang et al. suggested a charged cluster model in which nanometer-sized diamond clusters are formed in the gas phase by charge-induced nucleation.

Some calculations had also been made in their papers.

The "charged cluster" model could not explain why micron-size and even bigger diamond can be obtained by the activated CVD diamond process, because the so-called "capillary pressure" has already been lost during the diamond cluster growing up to micron-size. The basic concept of energy levels in that model must be incorrect as shown in Fig. 6.32, from which it can also be found that the transformation from so-called "charged diamond cluster" to diamond is more difficult than the transformation from graphite to diamond.[36~38]

Fig. 6.32 Energy level of "charged diamond clusters" and energy level of diamond should not be confused in "charged cluster model"

6.7.3 Under-Saturated Crystal Growth Model

In 1998 and 1999, W. Piekarczyk was still puzzled by activated CVD diamond processes.[45, 46] At the beginning of his two papers, "*How and why CVD diamond is formed: a solution of the thermodynamics paradox*" and "*Crystal growth of CVD diamond and some of its peculiarities*" he wrote that

> Diamond chemical vapor deposition (CVD) thought of as a crystal growth process, is a thermodynamic paradox, because it violates fundamental principles of thermodynamics. The most astonishing violation is the experimental observation that CVD diamond can form in gaseous environments that are carbon under-saturated with respect to diamond.

However, the premise does not exist, because no such "experimental observation" could be found in any experiments, except with reference to the limitation or prejudice of classical equilibrium thermodynamics. Of course, he hoped to overcome the limitation, so he continued to say in his paper that "it is shown that none of the principles is broken" when the peculiarities of CVD diamond are considered. Because of the influence of Van Rysselberghe's unsuitable speculation on a so-called common "coupling substance" in both the coupled reaction and coupling reaction, Piekarczyk did not accept our thermodynamic coupling model at first, although he later did list some conclusions from his model which were similar to ours. In fact, these two models are quite different.[36~38]

Besides these models mentioned above, there were still other thermodynamic discussions and models in the 1990s. For instance, I. Pinter, A. Tebano, A. Paoletti, P. Paroli; [47] N.A. Prijaya, J.C. Angus, P.K. Bachmann; [48] N.M. Hwang, J.H. Hahn, G.W. Bahng; [49] R. Bachmann, B. Sobisch, W. Kulich; [50] I.J. Ford, [51, 52] et al., also discussed or proposed models concerned with thermodynamics. Among them Rakov’s and Jacob’s papers were concerned with calculations of phase diagrams. [53, 54] However, Rakov said:

> calculations were based on the assumption that diamond cannot be deposited out of the graphite stability domain. Therefore, the lower concentration boundary of the graphite deposition region can be considered at the same time as the lower concentration boundary of the diamond deposition.

In fact, Rakov’s phase diagram was a typical graphite equilibrium phase diagram, calculated by classical equilibrium thermodynamics. Because the Gibbs free energies of diamond and graphite are close to each other, Rakov explained the activated low-pressure diamond growth by using graphite phase diagrams instead of real diamond phase diagrams. Another problem in Rakov’s phase diagram was the loss of the higher concentration boundary between the diamond deposition region and the nondiamond carbon deposition region. Jacob clearly recognized that his calculated diamond growth region was inside the graphite phase region. Therefore, no thermodynamic explanation can be found in Rakov’s and Jacob’s papers for the well-confirmed experimental observation of diamond deposition with the simultaneous graphite etching at activated low pressure.

6.8 “Nanothermodynamics” Model in 2005

In 2005, C.X. Wang and G.W. Yang (corresponding author) published a paper “*Thermodynamics of metastable phase nucleation at the nanoscale*”, and called it as “nanothermodynamics model”. [55] About that model they said:

> Our thermodynamic theory emphasizes the size effect of the surface tension of nanocrystals on the stable and metastable equilibrium states during the micro-phase growth.

and said:

> Firstly, we studied diamond nucleation upon chemical vapor deposition (CVD), and found out that the capillary effect of the nanosized curvature of diamond critical nuclei could drive the metastable phase region of the nucleation of CVD diamond into a new stable phase region in the carbon thermodynamic equilibrium diagram. Consequently, the diamond nucleation is preferable to the graphite phase formation in the competing growth of diamond and graphite upon CVD.

Such basic ideas were very similar to those of the charged cluster model, as mentioned-above in the last Section 6.7 , because in charged cluster model Hwang et al. in 1996, i.e. about nine years ago before 2005, said:

> When the carbon cluster is sufficiently small, the capillary pressure built up inside the cluster can be high enough to make diamond more stable than graphite.

and

> Hwang et al. suggested a charged cluster model in which nanometer-sized diamond clusters are formed in the gas phase by charge-induced nucleation.

said Hwang et al. in 1996, i.e. about nine years ago before 2005. Some quantitative calculation had also been made by Hwang et al. much earlier than that made by Yang et al. in 2005. Both of these two models emphasized "size effect or nanometer-sized" and "capillary effect or capillary pressure", which could "drive the metastable phase region of the nucleation of CVD diamond into a new stable phase region or make diamond more stable than graphite". However, both of these two models are incorrect as discussed in Section 6.7 for the charged cluster model. Figure 6.32 after a little revision, as shown in Fig. 6.33, could also be used for the "Nanothermodynamics model" here. Figure 6.33 also shows that if the so-called "nano-size diamond nuclei" were more stable than graphite, then the transformation from "nano-size diamond nuclei" to diamond crystals would be more difficult.

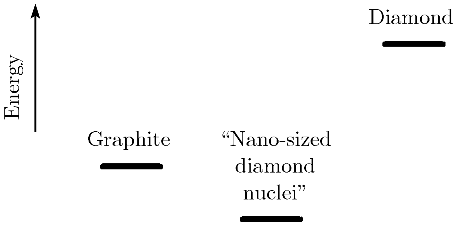


Fig. 6.33 Energy level of "nano-sized diamond nuclei" and energy level of diamond should not be confused in "nanothermodynamics" model

Moreover, some statements in Yang's 2005 paper were not true, such as [31]

> In thermodynamics aspects, several qualitative models have been proposed so far, respectively, i.e. Sommer's quasiequilibrium model, Yarbrough's surface reaction model, Bar-Yam's defect-induced stabilization model, Hwang's charged cluster model, and Wang's chemical pump model, and so on. However, none of these models clearly gives a completely satisfying insight into the formation of CVD diamonds, each model just tended to focus on one aspect of complicated process of the CVD diamond nucleation.

Here, first, Yang et al. did not talk accurately about all quantitative calculations of these models, as discussed in this chapter above, and proclaimed all of these models as being "qualitative". The next, Yang et al. incorrectly made a citation of comment [31] included in a paper published in the February 1990 issue of *Science* for all models even though for some models (i.e. Hwang's model and Wang's model) proposed after February 1990. Obviously, Yang et al. had juggled the original meanings of that comment.

In summary of this chapter, except for our thermodynamic coupling model for the activated low-pressure diamond growth, which will be discussed continuously in the following chapters, all other thermodynamic theoretical models or explanations did not break away from the limitation of classical thermodynamics.

References

1. Cui Y-G. Scientific Chinese [J], 1995(5): 48-50.
2. Bannister F A, Lonsdale K. Nature [J], 1943, 151: 334.

3. Rossini F D, Jessup R S. J Res Nat Bur Stand [J], 1938, 21: 491.
4. Berman R, Simon F. Z Electrochem [J], 1955, 59: 333.
5. Bromberg J P. Physical Chemistry [M]. 2nd edn. Boston: Allyn and Bacon Inc, 1984, 178-179.
6. Fu X-C, Shen W-X, Yao T-Y. Physical Chemistry [M]. 4th edn. Beijing: High Education Press, 1990, vol 1: 421.
7. Bundy F P, Hall H T, Strong H M, Wentorf R H. Nature [J], 1955, 176: 51-54.
8. Moore W J. Physical Chemistry [M]. 5th edn. London: Longman, 1976, Chap 6, 226.
9. Wentorl R H. General Electric R & D Center [R]. New York: 1970.
10. Deryagin B V, Fedoseev D V, Lukyanovich V M, Spitsyn B V, et al. J Cryst Growth [J], 1968, 2: 380-384.
11. Inst of Physical Chemistry, et al. Physicochemical Synthesis of Diamond from the Gas Phase [M]. Kiev: “Technik” Press, 1971, 18.
12. Deryagin B V, Spitsyn B V, Builovetal B V, et al. Dokl Akad Nauk [J], 1976, 231: 333.
13. Spitsyn B V, L. L. Bouilov and Deryagin BV. J. Cryst. Growth [J], 1981, 52: 219-226.
14. Spitsyn B V. The Roads to Metastable Diamond Growth. In: Diamond Materials, 3rd International Symposium [C], at Honolulu, Hawaii, May 16-21, 1993, ed. by Dismukas J P, Ravi K V, Spear K E, et al. Pennington, New Jersey: The Electrochemical Society Inc, 1993, PV93-17: 345-350.
15. Spitsyn B V. Chap. 10 Growth of Diamond Films from the Vapor Phase. In: Handbook of Crystal Growth [M]. ed by Hurle D T J. Amsterdam, NL: Elsevier Science, 1994, vol 1: 401-456.
16. Spitsyn BV. Crystallization of Diamond by the Chemical Transport Reaction: Thermodynamics and Kinetics. In: Diamond Materials, 4th International Symposium [C], at Reno, Nevada, May 21-26, 1995, ed by Ravi K V, Dismukas J P. Pennington, New Jersey: The Electro-chemical Society Inc, 1995, PV95-4: 61-72.
17. Matsumoto S, Sato Y, Tsutsumi M, Setaka N. J Materials Sci [J], 1982, 17: 3106-3112.
18. Kamo M, Sato Y, Matsumoto S, Setaka N. J Cryst Growth [J], 1983, 62: 642-644.
19. Setaka N. Proc of 3rd Internat Symp on Diamond Materials, Hawaii [C]. Pennington: The Electrochem Soc Inc, 1993, PV93-17: 1-8.
20. Matsumoto S, Sato Y, Kamo M, Setaka N. J Appl Phys [J], 1982, 21(4): L183-L185.
21. Matsumoto S, Hino M, Kobayashi. Appl Phys Lett [J], 1987, 51(10): 737-739.
22. Setaka N. J. Materials Research [J], 1989, 4(3): 664-670.
23. Spear K E. Earth & Mineral Sciences [J], 1987, 56(4): 53-56.
24. Jou S, Doerr H J, Bunshah R F, Thin Solid Films [J], 1994, 253: 95-102.
25. Angus J C, Science [J], 1988, 241: 913-921; Angus J C. Proc. of 1st International Symp on Diamond and Diamond-like Films [C], Pennington: The Electrochem Soc, 1989, PV89-12: 1-13; Angus J C. MRS Bulletin [R], October, 1989: 38-47.
26. Wang J-T, Cao C-B, Zhang P-J. J Electrochem Soc [J], 1994, 141(1): 278-281.
27. Sommer M, Mui K, Smith F W. Solod State Communication [J], 1989, 69(7): 775-778.
28. Batty J C, Stickney R R. J Chemical Physics [J], 1969, 51(10): 4475-4484.
29. Yarbrough W A. MRS Fall Meeting [C], Boston, 28 Nov 1989, Paper F1.3; Yarbrough W A. J Electronic Materials [J], 1991, 22(2): 133-139.
30. Bar-Yam Y, Moustakas T D. Nature [J], 1989, 342: 786.
31. Yarbrough W A, Missier R. Science [J], 1990, 247: 688-696.
32. Wang J-T, Carlsson J-O. Surf & Coat Tech [J], 1990, 43/44: 1-9; and presented at 8th Internat Conf on Thin Films [C]. San Diego, USA: 2–6 Apr 1990.

33. Wang J-T, Cao C-B, Zhang P-J. J Electrochem Soc [J], 1994, 141(1): 278-281.
34. Wang J-T, Zhang P-J, Cao C-B, et al. Proc. of 3rd Internat Symp on Diamond Materials [C], Honolulu, Hawaii: 21-26 May 1993. The Electrochem Soc Inc, 1993, PV93-17: 962-967.
35. Wang J-T, Zhang P-J. Supp Issue of Chinese J of Mater Res (I). Proc of Inter Union of Mater Res Soc, Internat Conf in Asia [C], 6-10 1993, Wuhan-Chongqing, ed. by Han Y-F, Huang Y, Tian J-M, Wang P-X (NSFC & C-MRS, 1993) 363-368.
36. Wang J-T. Modern thermodynamics – and a whole view of thermodynamics [M]. Shanghai: Fudan Univ. Press, 2005, 223 pages.
37. Wang J-T. Nonequilibrium Nondissipative Thermodynamics – With Application to Low-Pressure Diamond Synthesis [M]. Heidelberg: Springer, 2002.
38. Wang J-T. Phase Diagrams of Stationary Nonequilibrium States – Thermodynamics for Low Pressure Diamond Growth from the Vapor Phase [M]. Beijing: Science Press, 2000, 212 pages.
39. Wang J-T, Zhang DW, Liu Z-J. Thermodynamic Coupling Model for Low Pressure Diamond Growth from the Vapor Phase [M]. Beijing: Science Press, 1998, reprinted in 2000, 175 pages.
40. Gogotsi Y, Welz S, Ersoy DA, McNallan M.L. Nature [J], 2001, 411: 283.
41. Zhang Y-F, Zhang F-Q, Chen G-H. J Materials Research [J], 1994, 9: 2845-2849.
42. Wang J-T, Huang Z-Q, Wan Y-Z, Zhang D W, Jia H-Y. J Materials Research [J], 1997, 12(6): 1530-1535.
43. Hwang N M, Hahn J H, Yoon D Y. J Cryst Growth [J], 1996, 160: 87-97, 98-103; 1996, 162: 55-68; 1999, 204: 85-90. (1999)
44. Jiang H M, Hwang N M. J Materials Research [J], 1998, 13(12): 3536-3549.
45. Piekarczyk W. J Mater Sci [J], 1998, 33: 3443-3453.
46. Piekarczyk W. Cryst Res Technol [J], 1999, 34: 553-563.
47. Pinter I, Tebano A, Paoletti A, Paroli P. Diamond and Relat Mater [J], 1993, 3: 126-128.
48. Prijaya N A, Angus J C, Bachmann P K. Diamond and Relat Mater [J], 1993, 3: 129-136.
49. Hwang N M, Hahn J H, Bahng G W. Diamond and Relat Mater [J], 1993, 3: 163-167.
50. Bachmann R, Sobisch B, Kulich W. Diamond and Relat Mater [J], 1995, 4: 256-260.
51. Ford I J. J Appl Phys [J], 1995, 78(1): 510-513.
52. Ford I J. J Phys D: Appl Phys [J], 1996, 29: 2229-2234.
53. Rakov E G. Appl Phys Lett [J], 1996, 69(16): 2370-2372.
54. Jacob K T, Gundiah N, Menon A G. High Temperature Materials and Processes [J], 1996, 15(4): 223-235.
55. Wang C C, Yang G W. Materials Science and Engineering [J], 2005, R49: 157-202.

Chapter 7

Nondissipative Thermodynamics and Binary Nonequilibrium Phase Diagrams

Abstract As discussed in previous chapters, the extremum of state function is the necessary and sufficient condition for equilibrium. However, the equality of the second law had been practically used as the criterion of equilibrium for about 150 years since the establishment of thermodynamics. That is, only simple systems were under consideration in classical thermodynamics. Now it is known that $[\mathrm{d_i}S_1 < 0, \mathrm{d_i}S_2 > 0 \& \mathrm{d_i}S = 0]$ belongs to the equality part of the second law, but it is a new field of nonequilibrium nondissipative thermodynamics which provides a theoretical base for nonequilibrium phase diagrams. There is at least one more dimension of nonequilibrium conditions for nonequilibrium phase diagrams in comparison with the classical or traditional equilibrium phase diagrams. Theoretically calculated nonequilibrium phase diagrams agree with experimental data of activated low-pressure diamond synthesis. The calculation principle and method of nonequilibrium phase diagrams are discussed in detail in this Chapter.

7.1 An ABC in Mathematics

Mathematical functions are multifarious and numerous. Among them some kinds of mathematical functions are of extremum inside the discussed regions, and some are not of extremum. If the first order derivative (or differential) of a mathematical function equals to zero and its second order derivative (or differential) is bigger than zero, then there is a minimum at the corresponding point, as shown in Fig. 7.1 for the function of $y = x^2$.

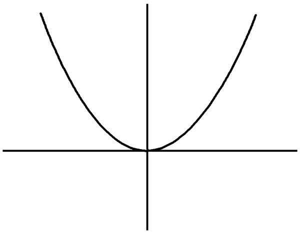


Fig. 7.1 A minimum at $x = 0$ for a function ($y = x^2, y' = 2x$ and $y" = 2 > 0$)

If the first order derivative (or differential) of a mathematical function equals to zero and its second order derivative (or differential) is smaller than zero, then there is a maximum at the corresponding point, as shown in Fig. 7.2 for the function of $y = -x^2$. If the first order derivative (or differential) of a mathematical function equals to zero and its second order derivative (or differential) is also equal to zero, then there may be no extremum at the corresponding point, as shown in Fig. 7.3(a) and Fig. 7.3(b) for the functions of $y = x^3$ and $y = -x^3$. Therefore, we should not merely, on the basis of zero value of the first order derivative (or differential) of a mathematical function, assert the extremum or the point of extremum inside the discussed region. We must make a further examination of its second order derivative (or differential).[1]

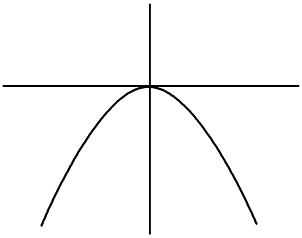
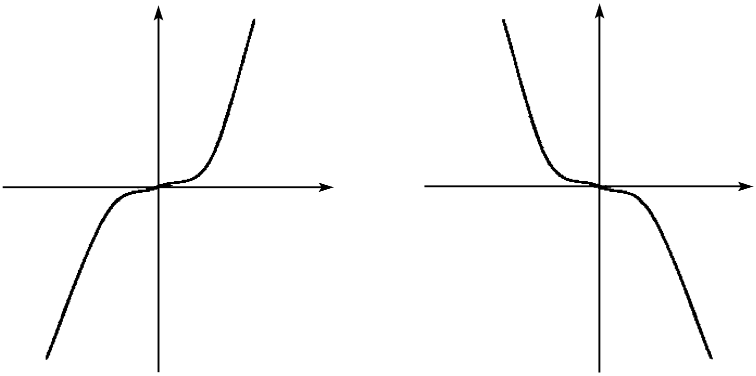


Fig. 7.2 A maximum at $x = 0$ for a function ($y = -x^2, y' = -2x$ and $y" = -2 > 0$)



(a) $y = x3, y' = 3x^2$ and $y'' = 6x$ (b) $y = -x^3, y' = -3x^2$ and $y'' = -6x$

Fig. 7.3 Two kinds of mathematical functions without extremum at $x = 0$.
($y = x^2$, $y' = 2x$ and $y'' = 2 > 0$)

In classical thermodynamics, the research systems are equilibrium systems or systems including a simple irreversible process from an initial equilibrium state to another final equilibrium state. It will be better to say, the research systems are uncoupling systems in classical thermodynamics. For a system under isothermal and isobaric conditions, its state function of Gibbs free energy is of a minimum, similar to that as shown in Fig. 7.1. If its first order derivative (or differential) equals to zero and its second order derivative (or differential) is bigger than zero, then the system is at a minimum of its state function of Gibbs free energy, and the system is in its equilibrium state, similar to that the system is at the point of $x = 0$ in Fig. 7.1. For an isolated system its state function of entropy is of a maximum, similar to that as shown in Fig. 7.2. If its first order derivative (or differential) equals to zero and its second order derivative (or differential) is smaller than zero, then the system is at a maximum of its state function of entropy, and the system is in its equilibrium state, similar to that the system is at the point of $x = 0$ in Fig. 7.2.

For complex multi-irreversible-process systems, the corresponding state functions may include only terrace points but no extremum inside the discussion or interesting region, as shown in Fig. 7.3(a) and in Fig. 7.3(b). In such cases, even if the first order derivative (or differential) of the corresponding state function equals to zero, the system may be not in equilibrium. If the second order derivative (or differential) of the corresponding state function also equals to zero, the system may be in a nonequilibrium state, but not in an equilibrium state.[1~8]

In fact, for complex multi-process systems the number of variables increases, and the change of the state function may not be presented by ichnography in a two-dimensional coordinate but by curving surface in a three-dimensional (or multidimensional) coordinate. In such cases, saddle points should also be under consideration, as shown in Fig. 7.4. The first order partial differential of the saddle point equals to zero, but there is no extremum. Therefore, the first order partial differential equal to zero is not a sufficient condition to ensure generally the system being in equilibrium.

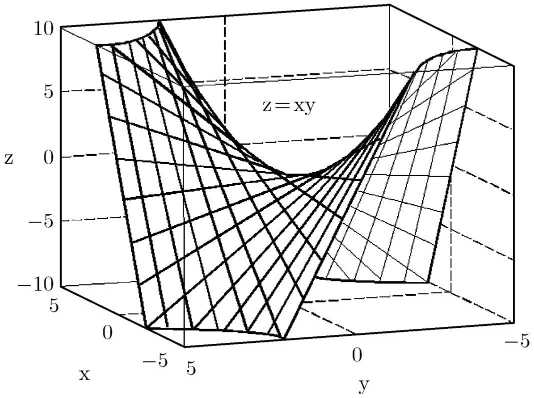


Fig. 7.4 A curving surface function with a terrace point at $x = 0$ and $y = 0$

Some thermodynamics textbooks in discussion of the criterion of equilibrium elucidate: (p. 152)[9]

> From the second law of thermodynamics, i.e., a theory concerned with irreversible processes, we may obtain a necessary and sufficient condition to determine the possibility for changing a system from a state to another state. For instance, if the system under consideration is changing adiabatically, any change will be toward entropy increase of the system. ··· Therefore, if the system changes adiabatically, the necessary and sufficient condition for the system in thermodynamic equilibrium is a maximum of entropy.

Moreover, (p. 154)[9], (p. 158)[9]

> ··· in mathematics, it had been demonstrated that during a function at an extremum, in spite of maximum or minimum, the first order differential of the function must equal zero. ···
>
> The conditions mentioned-above for all of thermal equilibrium, mechanical equilibrium and phase equilibrium are obtained from that the first order differential of some thermodynamic functions equals zero. These conditions are merely the necessary conditions, but not sufficient to determine the realization of equilibrium. For getting the sufficient condition the discussion on the second order differential of the concerning thermodynamic function should be made. ···

and so on. In some textbooks, it was very clearly outlined that the criterion (i.e. the necessary and sufficient condition) of equilibrium for isolated systems is $\delta S = 0$ together with

$\delta^2 S < 0$, while the criterion of equilibrium for isothermal isobaric systems is $\delta G = 0$ together with $\delta^2 G > 0$. Here, δ represents any kind of possible virtual alteration, because inside equilibrium system any actual macroscopic alteration does not exist.[10]

However, in most thermodynamics textbooks the equality of the mathematical expression of the second law of thermodynamics was practically used as the criterion of equilibrium, or as the criterion of reversible process for the discussed process. That is, the equality of the mathematical expression of the second law of thermodynamics $[\mathrm{d_i}S = 0]$ had directly corresponded to equilibrium systems or reversible processes, while the inequality of the mathematical expression of the $[\mathrm{d_i}S > 0]$ had directly corresponded to nonequilibrium systems or irreversible processes. In fact, all of these statements are suitable only for discussing simple uncoupling systems in classical thermodynamics, or only under assumptions which had been pointed out previously. It will be clearer to use a clear mathematical expression of the second law of classical thermodynamics, i.e. $[\mathrm{d_i}S_\mathrm{p} = \mathrm{d_i}S \geqslant 0]$. Then a strict statement should be: the equality $[\mathrm{d_i}S_\mathrm{p} = \mathrm{d_i}S = 0]$ corresponds to equilibrium system or reversible process, while the inequality $\mathrm{d_i}S_\mathrm{p} = \mathrm{d_i}S > 0]$ corresponds to nonequilibrium system or irreversible process. Otherwise, we should say: equilibrium system or reversible process must belong to the equality part of the mathematical expression of the second law of thermodynamics, but not vice versa. It will better to say that nondissipative systems must correspond to the equality part of the mathematical expression of the second law of thermodynamics, and dissipative systems must be corresponding to the inequality part of the mathematical expression of the second law of thermodynamics. Such a statement can be used both for simple uncoupling systems (including equilibrium systems and systems including only spontaneous processes) and complex multi-process coupling systems. Following a correct rigorous criterion (i.e. a necessary and sufficient condition) and the local equilibrium approximation, the emergences of a new field of nondissipative thermodynamics and the modern classification of thermodynamics are inevitable without any assumptions. The reason for such a problem being hidden for such a long time might be that no clear opposite examples had been found. Therefore, the emergence of the activated low-pressure diamond synthesis and some other new technologies became the necessary and objective historical condition for the establishment of nondissipative thermodynamics.

Meanwhile, such simplistic mathematics may also be a cause for another puzzle existing for a long time in the field of thermodynamics. It was mentioned in Chapter 4, Section 4.8.2 Traditional Classifications of Thermodynamics, that "linear nonequilibrium thermodynamics" and "close-to-equilibrium nonequilibrium thermodynamics" are different concepts, and they should not be treated as synonyms. After Onsager and Prigogine won their Nobel Prizes in 1968 and 1977, respectively, the confusion between "linear nonequilibrium state" and "close-to-equilibrium nonequilibrium state" could often be found in the classification of thermodynamics. Correspondingly, the confusion between "nonlinear nonequilibrium state" and "far-from-equilibrium nonequilibrium state" could also be found. Now, based on a few figures of mathematical curves, these different concepts can be clearly distinguished. For instance, the state near $x = 0$ in Fig. 7.3 may belong to linear region of thermodynamic state function, but not close-to-equilibrium region of thermodynamic state function, while the states near $x = 0$ in Fig. 7.1 or Fig. 7.2 really belong to close-to-equilibrium region of thermodynamic state function. Simply speaking, the state "close-to-equilibrium" must be located near the extremum of thermodynamic state function, but there is no such limitation for "linear nonequilibrium state". The term of "nearly in equilibrium" and "assumption of microscopic reversibility" or assumption of "detailed balancing" were used in Onsager's paper; so Onsager reciprocal relations might be used in "nearly in equilibrium" region, but not as a strict thermodynamic fundamental law in the field of "linear nonequilibrium thermodynamics".[11] Of course, Fig. 7.1 – Fig. 7.3 are figures in low dimensional space; there may be more complicated situations in multiple dimensional space.

7.2 The Nature of CALPHAD

It is well-known that calculation of phase diagrams (CALPHAD) is an important means of connection between thermodynamics and practical applications. Phase diagrams can clearly show the results of thermodynamic calculations. Different phase regions represent the appearance or existence of some special phases, but not of other phases. The phase line lies between two neighbor phase regions. The places of the phase line represent the critical conditions for the coexistence of these two neighbor phases. Figure 7.5 illustrates the simplest equilibrium phase diagram of classical thermodynamics without external restricted conditions.

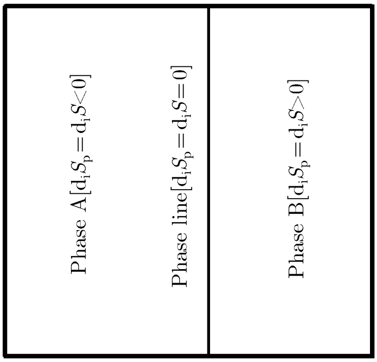


Fig. 7.5 An illustration of the simplest equilibrium phase diagram

Here, d_iS_p and d_iS are entropy production of process(es) and entropy production of the uncoupling system, respectively. If Fig. 7.5 shows a transformation A = B (such as the phase transformation from ice to water), the vertical coordinate may be pressure p, while the horizontal coordinate may be temperature T. Under the conditions inside the left phase region the transformation of A = B would be a nonspontaneous process, and thermodynamically impossible to take place alone. That is, both entropy production of process(es) and entropy production of the uncoupling system would be smaller than zero [$d_iS_p = d_iS < 0$]. Therefore, only phase A exists inside the left region of Fig. 7.5, and the left side of the phase diagram is called phase A region. On the other hand, under the conditions inside the right phase region, the transformation of A = B is a spontaneous process, i.e. [$d_iS_p = d_iS > 0$]. Therefore, only phase B exists inside the right region of Fig. 7.5, and the right side of the phase diagram is called phase B region. The boundary at the middle is the phase line. Under the conditions on the line, [$\mathrm{d_i}S_\mathrm{p} = \mathrm{d_i}S = 0$]. That is the critical condition for the coexistence of A and B.

If the research system is a multi-process system, in which several processes occur simultaneously, and $\mathrm{d_i}S_2 > 0$ is a constrained condition keeping constantly and never disappear during the whole research, then the corresponding phase diagram may be as shown in Fig. 7.6. That is a nonequilibrium phase diagram. In the activated low-pressure CVD diamond growth, $\mathrm{d_i}S_2 > 0$ corresponds to the association of superequilibrium atomic hydrogen. During the whole process of the activated low-pressure CVD diamond growth the reaction rate keeps constantly. The system is a nonequilibrium system and the corresponding phase diagram is, of course, a nonequilibrium phase diagram.

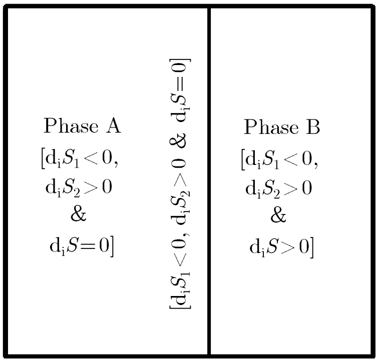


Fig. 7.6 An illustration of the simplest nonequilibrium phase diagram

From Fig. 7.5 and Fig. 7.6 it can be found that a phase diagram is an illustration of three cases [$\mathrm{d_i}S < 0$], [$\mathrm{d_i}S = 0$] and [$\mathrm{d_i}S > 0$] of the second law of thermodynamics, in principle. There is nothing concerned with the second order differential. Therefore, calculations of phase diagrams (CALPHAD) could not be limited only for equilibrium phase diagrams. Traditionally it was believed that phase diagrams must be equilibrium phase diagrams. However, that was merely the limitation of classical thermodynamics in cognition. The nature of phase diagrams is diagram without dissipation, so phase diagrams may be equilibrium nondissipative diagrams, i.e. equilibrium phase diagrams, and may be nonequilibrium nondissipative diagrams, i.e. nonequilibrium phase diagrams. [11~15]

Equilibrium phase diagrams are independent of time, so they may be illustrated in books or articles. Nonequilibrium phase diagrams may change with time, so it will be better to illustrate nonequilibrium phase diagrams by motion video. For convenience of research and discussion only the stationary nonequilibrium phase diagrams, which are independent of time, will be illustrated in this book. The activated low-pressure diamond synthesis is usually running under stationary nonequilibrium conditions for the diamond growth, so the discussion on stationary nonequilibrium phase diagrams is favorable for not only the theoretical guide to the new process, but also the comparison between theory and experiments. In this bok phase diagrams noted by nonequilibrium phase diagrams are, of course, stationary nonequilibrium phase diagrams independent of time, because all nonequilibrium phase diagrams printed on papers must be independent of time.

Be careful, the calculations of nonequilibrium phase diagrams are different from the calculations of equilibrium phase diagrams. For equilibrium phase diagrams, if the composition, temperature, and pressure are fixed, the information of phases is calculated only based on $[\mathrm{d_i}S_1 = 0, \mathrm{d_i}S_2 = 0 \& \mathrm{d_i}S = 0]$ or $[\mathrm{d_i}S_\mathrm{p} = \mathrm{d_i}S = 0]$. However, for nonequilibrium phase diagrams calculations, the fixed values of composition, temperature, and pressure are not sufficient, and we must know the constrained experimental conditions designed externally, i.e. the intensity of the constrained experimental conditions, or the absolute value of $\mathrm{d_i}S_2$ ($\mathrm{d_i}S_2 > 0$) during the unit interval of time. In this way, one can calculate the situations of phases on the basis of nonequilibrium nondissipation principle, i.e. $[\mathrm{d_i}S_1 < 0, \mathrm{d_i}S_2 > 0 \& \mathrm{d_i}S = 0]$. For nonequilibrium phase diagram calculation, at least, one more variable is needed than that for equilibrium phase diagram calculation. This is the obvious distinction between nonequilibrium phase diagram calculation and equilibrium phase diagrams calculation. If $|\mathrm{d}S_2| = 0$ and, of course, $|\mathrm{d}S_1| = 0$, then the nonequilibrium phase diagram calculation will be degraded into the equilibrium phase diagram calculation on the basis of $[\mathrm{d_i}S_1 = 0, \mathrm{d_i}S_2 = 0 \& \mathrm{d_i}S = 0]$ or $[\mathrm{d_i}S_\mathrm{p} = \mathrm{d_i}S = 0]$. The corresponding nonequilibrium phase diagrams will be degraded into the equilibrium phase diagrams. Therefore, classical thermodynamics is a special case of modern thermodynamics; or better to say, modern thermodynamics is an extension of classical thermodynamics, and nonequilibrium phase diagrams are established by extension of equilibrium phase diagrams. Of course, the term, “modern thermodynamics”, here has an extended meaning of thermodynamics at the present time for the whole thermodynamics discipline.

If the value of $|\mathrm{d}S_2| \neq 0$ during the unit interval of time, then the system must be a nonequilibrium system. For the activated low-pressure synthesis, the practical or designed value of the concentration of superequilibrium atomic hydrogen could be known. There are only two main irreversible processes included inside the new activated low-pressure diamond synthesis, so the nonequilibrium parameter needed for characterization of the system should be only one. The skill for nonequilibrium phase diagram calculation is also important. The following part of this chapter will continue with a systematic description on nondissipative thermodynamics and nonequilibrium phase diagram calculations.

7.3 Nondissipative Thermodynamics and Nonequilibrium Phase Diagrams

As mentioned above, the criterion of nondissipative thermodynamics (or what can be called nonequilibrium nondissipative thermodynamics, or modern nondissipative thermodynamics) is $[\mathrm{d_i}S_1 < 0, \mathrm{d_i}S_2 > 0 \& \mathrm{d_i}S = 0]$ or $[(\mathrm{d}G_1)_{T,p} > 0, (\mathrm{d}G_2)_{T,p} < 0 \& (\mathrm{d}G)_{T,p} = 0]$ for isothermal isobaric systems. Nondissipative thermodynamics belongs to the equality part of mathematical expression of the second law of thermodynamics, so quantitative calculation can be made in the new field. Due to the existence of constrained condition, $\mathrm{d_i}S_2 > 0$ or $(\mathrm{d}G_2)_{T,p} < 0$, the system will not be able to reach equilibrium, but stationary nonequilibrium state instead. Therefore, the illustration of phase situations of calculation results is just the stationary nonequilibrium phase diagram, or known directly as nonequilibrium phase diagram. Calculation of nonequilibrium phase diagram is somewhat different from that of equilibrium phase diagram. The calculation of equilibrium phase diagram is based on Gibbs free energy minimization principle for isothermal isobaric systems, while the calculation of nonequilibrium phase diagram should be based on nonequilibrium nondissipation principle, i.e. $[(\mathrm{d}G_1)_{T,p} > 0,\ (\mathrm{d}G_2)_{T,p} < 0\&\,(\mathrm{d}G)_{T,p} = 0]$ or $[\mathrm{d_i}S_1 < 0,\mathrm{d_i}S_2 > 0\&\mathrm{d_i}S = 0]$. Nonequilibrium nondissipative thermodynamics is a new field, so a rather systematic discussion on nondissipative thermodynamics and calculation of nonequilibrium phase diagrams will therefore be made in this chapter, and the activated low-pressure diamond growth will be taken as an example for detailed discussions and quantitative calculations. [16~50]

As there may be only two elements (carbon and hydrogen) included in the activated low-pressure diamond growth system, and as all the corresponding thermodynamic data can be found, quantitative calculations can be made without assumption or estimation on thermodynamic data. However, the method and principle should not be on the basis of classical equilibrium thermodynamics, and should not directly use the ordinary commercial programs for calculation of phase diagrams, which are usually on the basis of Gibbs free energy minimization principle. If the total reaction Eq. (**??**) of thermodynamic coupling model in Chapter 6 is calculated on the basis of Gibbs free energy minimization principle, then superequilibrium atomic hydrogen will automatically change into atomic hydrogen of equilibrium concentration. That will result in producing graphite but not diamond. That is, under low pressure equilibrium conditions the transformation from graphite to diamond is thermodynamically impossible. However, such a conclusion is meaningless for the activated low-pressure diamond growth, because every practical experimental researcher knows that the stable low-pressure diamond growth must be under nonequilibrium conditions.

To keep the concentration of superequilibrium atomic hydrogen constant during calculation, the basic reaction equations of thermodynamic coupling model, i.e. Eq. (6.12), Eq. (6.13), Eq. (6.14) or Eq. (6.15), Eq. (6.16), Eq. (6.17), could be rewritten into the following forms:

$$\begin{gathered}\mathrm{C(gra)} = \mathrm{C(dia)}, \\ (T, p \leqslant 10^5\ \mathrm{Pa}) \\ (\Delta G_1)_{T,p} > 0, \Delta_\mathrm{i} S_1 < 0.\end{gathered} \tag{7.1}$$

$$\begin{gathered}\mathrm{H}* = 0.5\mathrm{H}_2, \\ (T_\mathrm{activation} \gg T, p \leqslant 10^5\ \mathrm{Pa}) \\ (\Delta G_2)_{T,p} \ll 0, \Delta_\mathrm{i} S_2 \gg 0.\end{gathered} \tag{7.2}$$

The total reaction is

$$\begin{gathered}\mathrm{C(gra*)} = \mathrm{C(dia)}, \\ (T_\mathrm{activation} \gg T, p \leqslant 10^5\ \mathrm{Pa}) \\ (\Delta G)_{T,p} \leqslant 0, \Delta_\mathrm{i} S \geqslant 0.\end{gathered} \tag{7.3}$$

Here,

$$\mathrm{C(gra*)} = [\mathrm{C(gra)} + \chi\ (\mathrm{H*} - 0.5\ \mathrm{H_2})] \tag{7.4}$$

and C (gra*) is called activated graphite. Obviously, thermodynamic data of activated graphite could not be found in thermodynamic databooks. Thermodynamic data of activated graphite must be related with concentration of superequilibrium atomic hydrogen (H*) and the detailed working conditions. That is an important feature in calculation of nonequilibrium phase diagrams. In calculation of equilibrium phase diagram there is only one equilibrium phase diagram for system with the same composition, while in calculation of nonequilibrium phase diagram there may be several nonequilibrium phase diagrams for system with the same composition but with different activation (or nonequilibrium) intensity. That is, nonequilibrium phase diagram has at least one more dimension for nonequilibrium condition than equilibrium phase diagram. For instance, for the activated low-pressure diamond synthesis there must be different calculation results for different hot filament temperatures, 2400 K or 2300 K. The changes of results must be continuous. Thermodynamic data of activated graphite will be calculated in the first step, and other steps for calculation of nonequilibrium phase diagrams will be described one by one as follows.

7.4 Thermodynamic Data of Activated Graphite

As thermodynamic data of superequilibrium atomic hydrogen and the corresponding thermodynamic data of activated graphite could not be found in standard thermodynamic databooks, we must find a way to deduce them. First of all, a clear definition of superequilibrium atomic hydrogen should be given: superequilibrium atomic hydrogen is defined in this book as atomic hydrogen of equilibrium hydrogen atom concentration at activation temperature $T_{\text{activation}}$ (such as the hot filament temperature) in the hot filament process, so for the substrate temperature T atomic hydrogen is of superequilibrium concentration. The derivation of thermodynamic data for superequilibrium atomic hydrogen and activated graphite will be discussed in what follows. In brief, the derivations here are limited only to $T_{\text{activation}} = 2400$ K, but in two different ways.

7.4.1 Gibbs Free Energy Method

1. If the substrate temperature equals the activation temperature, the whole system may be treated as an equilibrium system, so at 2400 K the Gibbs free energy of the activated graphite $G^{2400}_{\text{gra}*}$ should equal the Gibbs free energy of equilibrium graphite G^{2400}_{gra}, i.e.

$$G^{2400}_{\text{gra}*} = \left[G^{2400}_{\text{gra}} + \chi \left(G^{2400}_{\text{H}*} - 0.5\, G^{2400}_{\text{H}_2}\right)\right] = G^{2400}_{\text{gra}} \tag{7.5}$$

That is,

$$G^{2400}_{\text{H}*} = 0.5\, G^{2400}_{\text{H}_2}. \tag{7.6}$$

From thermodynamic databooks,[51] the standard Gibbs free energy of molecular hydrogen at 2400 K is

$$G_{H_2}^{2400} = -400.290\text{kJmol}^{-1}. \tag{7.7}$$

Therefore, the Gibbs free energy of superequilibrium atomic hydrogen at 2400 K is

$$G_{H*}^{2400} = -200.145.\text{kJmol}^{-1}. \tag{7.8}$$

For comparison, the standard Gibbs free energy of atomic hydrogen in thermodynamic databooks is

$$G_{H}^{2400} = -117.429\text{kJmol}^{-1} \tag{7.9}$$

In fact, the difference between these free energies is due to the difference of the partial pressure of superequilibrium atomic hydrogen from the standard pressure.

2. From thermodynamics, $G = H - TS$. For an ideal gas, the values of Gibbs free energy G and entropy S are closely related to the partial pressure (or concentration), but the value of enthalpy H is not. Approximately, $H_{H*} = H_H$, so

$$S_{H*} = S_H - (G_{H*} - G_H)/T, \tag{7.10}$$

$$\begin{aligned} S_{H*}^{2400} &= \left[S_H^{2400} + \left(\frac{200145 - 117429}{2400}\right)\right] \text{Jmol}^{-1}\text{K}^{-1} \\ &= \left[S_H^{2400} + 34.465\right] \text{Jmol}^{-1}\text{K}^{-1}. \end{aligned} \tag{7.11}$$

The heat capacity of superequilibrium atomic hydrogen at constant pressure C_p is taken to be the same as that of atomic hydrogen in the equilibrium state. Therefore, the equality is valid for the whole range of interest, i.e.

$$S_{H*}^{2400} = S_{H*}^{298} + \int_{298}^{2400} \frac{C_p}{T} \mathrm{d}T \tag{7.12}$$

and

$$S_{H}^{2400} = S_{H}^{298} + \int_{298}^{2400} \frac{C_p}{T} \mathrm{d}T, \tag{7.13}$$

so

$$\begin{aligned} S_{H*}^{298} &= \left[S_H^{298} + 34.465\right] \text{Jmol}^{-1}\text{K}^{-1} \\ &= [114.609 + 34.465]\text{Jmol}^{-1}\text{K}^{-1} \\ &= 149.073\text{Jmol}^{-1}\text{K}^{-1}. \end{aligned} \tag{7.14}$$

3. According to these data of superequilibrium atomic hydrogen and expression Eq. (7.4) for activated graphite, thermodynamic data for activated graphite with different χ values can be obtained, as listed in Table 7.1.

Note that in Eq. (7.7) and Eq. (7.9) standard thermodynamic data, i.e. data at standard pressure (101.325 kPa or 1 atmosphere), are used. If thermodynamic data at different pressures are substituted into Eq. (7.7) and Eq. (7.9), then the corresponding thermodynamic data for activated graphite at different pressures can be obtained. However, in fact, after the calculation it can be found that the data listed in Table 7.1 can be used without obvious error in the pressure range of interest in this book (0.010 kPa–1 × 101.325 kPa).

Table 7.1 Thermodynamic data of activated graphite (gra*) with different χ values

$H_{\mathrm{gra}*} = 217.9861\ \chi$ (kJ mol^{-1})			
$S_{\mathrm{gra}*} = 5.6861 + 83.7815\ \chi$ (J mol^{-1} K^{-1})			
C_p (J mol^{-1} K^{-1}) $= A + B \times 10^{-3}T + C \times 10^{5}T^{-2} + D \times 10^{-6}T^{2}$			
	298–1000 K	1000–1550 K	1550–2400 K
A	$0.1597 + 5.7635\chi$	$25.5701 + 5.7635\chi$	$25.5701 + 5.9518\chi$
B	$38.8990 + 1.2400\chi$	$-0.2232 + 1.2400\chi$	$-0.2232 - 1.6351\chi$
C	$-1.5154 + 0.2988\chi$	$-39.1476 + 0.2988\chi$	$-39.1476 + 20.6857\chi$
D	$-17.4121 - 1.3416\chi$	$0.0978 - 1.3416\chi$	$0.0978 + 0.1076\chi$

Activation temperature: 2400 K, Ref. Temp.: 298 K, Pressure range: 0.010–100 kPa

Definitions of B, C and D here are a little different from those in text

7.4.2 Equilibrium Constant Method

For superequilibrium atomic hydrogen and activated graphite data, a more careful derivation from equilibrium constants can be introduced. In the derivation in the latter paragraphs an approximation of the partial pressure of molecular hydrogen equal to the total pressure had been used. In fact, if the total pressure is lower and the activation temperature is higher, then an obvious deviation may be introduced. To remove this deviation, a derivation can be made by an equilibrium constant method.

1. The basic data of atomic hydrogen and molecular hydrogen at standard temperature and pressure can be found from thermodynamic databooks. According to the relationship G–T of Eq. (3.74) in Chapter 3, i.e.

$$G = H_0 - AT\ln T - \frac{1}{2}BT^2 - \frac{1}{2}CT^{-1} - \frac{1}{6}DT^3 + IT,$$

the corresponding G^0 values and the ΔG^0 value of the equilibrium reaction H $= 0.5$ H_2 at standard pressure and activated temperature can be obtained. Based on $\Delta G^0_{\mathrm{m}} = -RT\ln K^{\mathrm{o}}_p$ or $\Delta G^0 = -nRT\ln K^{\mathrm{o}}_p$, i.e. Eq. (3.112), nondimensional thermodynamic reaction equilibrium constants K^{o}_p can be obtained. If all the reactants and products are treated as ideal gases, then

$$K^{\mathrm{o}}_p = \left(\frac{p^e_{\mathrm{H_2}}}{p^0}\right)^{0.5}\left(\frac{p^e_{\mathrm{H}}}{p^0}\right)^{-1}. \tag{7.15}$$

Here, p^e represents the equilibrium partial pressure. K^{o}_p is a function of temperature and independent of pressure. When the reaction occurs at 2400 K, K^{o}_p (2400 K) = 63.130. Therefore, the equilibrium partial pressure of atomic hydrogen p^e_{H} (2400 K) and the equilibrium partial pressure of molecular hydrogen $p^e_{\mathrm{H_2}}$ can be calculated, as shown in Table 7.2.

Table 7.2 Equilibrium partial pressures of atomic hydrogen p_{H}^{e} and molecular hydrogen $p_{\mathrm{H}_2}^{e}$ at different total pressures p
(Activation temperature: 2400 K)

p	p_{H}^{e} (2400 K)	$p_{\mathrm{H}_2}^{e}$ (2400 K)
1 × 101.325 kPa	0.015 715 × 101.325 kPa	0.984 28 × 101.325 kPa
0.1 × 101.325 kPa	0.004 885 3 × 101.325 kPa	0.095 115 × 101.325 kPa
0.01 × 101.325 kPa	0.001 463 5 × 101.325 kPa	0.008 536 5 × 101.325 kPa
0.000 1 × 101.325 kPa	0.000 076 61 × 101.325 kPa	0.000 023 39 × 101.325 kPa

2. From the definition of superequilibrium atomic hydrogen (H*) above, it can be seen that $p_{\mathrm{H}*}$ = p_{H}^{e} (2400 K). The relationship between G and the pressures is Eq. (**??**) in Chapter 3, i.e. $G_{\mathrm{m}}(p,T) = G_{\mathrm{m}}^{0}\left(p^{0},T\right) + RT\ln\left(\frac{p}{p^{0}}\right)$. According to the values of $p_{\mathrm{H}*}$ and $p_{\mathrm{H}_2}^{e}$ (2400 K) the following equations can be obtained:

$$G_{\mathrm{H}*}^{298} = G_{\mathrm{H}}^{0,298} + R \times 298.15 \times \ln\left(\frac{p_{\mathrm{H}*}^{e}}{p^{0}}\right) \tag{7.16}$$

and

$$G_{\mathrm{H}_2}^{e,298} = G_{\mathrm{H}_2}^{0,298} + R \times 298.15 \times \ln\left(\frac{p_{\mathrm{H}*}^{e}}{p^{0}}\right). \tag{7.17}$$

Substituting them for the corresponding values in $S_{\mathrm{H}*}^{298} = (H_{\mathrm{H}*}^{298} - G_{\mathrm{H}*}^{298})/(298.15)$, and $S_{\mathrm{H}_2}^{e,298} = (H_{\mathrm{H}_2}^{298} - G_{\mathrm{H}_2}^{e,298})/(298.15)$, entropy of superequilibrium atomic hydrogen $S_{\mathrm{H}*}^{298}$ and entropy of molecular hydrogen $S_{\mathrm{H}_2}^{e,298}$ at the filament temperature (2400 K) and different total pressures p, can be obtained, respectively, as shown in Table 7.3. Here, the ideal gas assumption is also adopted. The value of the enthalpy H is independent of the partial pressure (or concentration), so $H_{\mathrm{H}*}^{298} = H_{\mathrm{H}}^{298}$.

Table 7.3 Entropy of superequilibrium atomic hydrogen $S_{\mathrm{H}*}^{298}$ and Entropy of molecular hydrogen $S_{\mathrm{H}_2}^{e,298}$ at different total pressures p
(Activation temperature: 2400 K)

p	$S_{\mathrm{H}*}^{298}$	$S_{\mathrm{H}_2}^{e,298}$
1 × 101.325 kPa	149.14 J K^{-1} mol^{-1}	130.71 J K^{-1} mol^{-1}
0.1 × 101.325 kPa	158.84 J K^{-1} mol^{-1}	150.14 J K^{-1} mol^{-1}
0.01 × 101.325 kPa	168.88 J K^{-1} mol^{-1}	170.19 J K^{-1} mol^{-1}
0.000 1 × 101.325 kPa	193.40 J K^{-1} mol^{-1}	219.24 J K^{-1} mol^{-1}

3. It is known that the standard entropy of graphite at 298 K is S_{gra}^{298} = 5.6860 J mol^{-1} K^{-1}. According to Eq. (7.4), the definition of activated graphite is C(gra*) = [C(gra) + χ (H*− 0.5 H_2)], and entropy of activated graphite is $S_{\mathrm{gra}*} = \left[S_{\mathrm{gra}} + \chi\left(S_{\mathrm{H}*} - 0.5S_{\mathrm{H2}}\right)\right]$. Entropy of activated graphite $S_{\mathrm{gra}*}$ can be obtained at different total pressures with χ = 0.28, as shown in Table 7.4. Obviously, in the pressure range of interest (0.0001 × 101.325 kPa – 1 × 101.325 kPa), entropy of activated graphite may be treated as a constant value independent of pressure. Based on Eq. (3.74) $G = H_0 - AT\ln T -$

$\frac{1}{2}BT^2 - \frac{1}{2}CT^{-1} - \frac{1}{6}DT^3 + IT$ and $G = H - TS$, Gibbs free energy of activated graphite can be obtained at different temperatures

Table 7.4 Entropy of activated graphite $S_{\text{gra}*}$ at different total pressures p (Activation temperature: 2400 K)

1 × 101.325 kPa	29.146 J K^{-1} mol^{-1}
0.1 × 101.325 kPa	29.144 J K^{-1} mol^{-1}
0.01 × 101.325 kPa	29.146 J K^{-1} mol^{-1}
0.000 1 × 101.325 kPa	29.144 J K^{-1} mol^{-1}

4. The enthalpy and the heat capacity of superequilibrium atomic hydrogen have already been treated by the ideal gas approximation, i.e. $H_{\text{H}*} = H_{\text{H}}$ and $C_{p\text{H}*} = C_{p\text{H}}$. Finally, the calculated results are listed in Table 7.5. Definitions of B, C and D in tables are a little different from those in the text.

Table 7.5 Thermodynamic data for activated graphite (gra*) ($\chi = 0.28$)

$H_{\text{gra}*} = 61.036$ (kJ mol^{-1})

$S_{\text{gra}*} = 29.145$ (J mol^{-1} K^{-1})

C_p (J mol^{-1} K^{-1}) $= A + B \times 10^{-3}T + C \times 10^5 T^{-2} + D \times 10^{-6} T^2$

	298–1000 K	1000–1550 K	1550–2400 K
A	1.7735	27.184	27.2366
B	39.236	0.1240	−0.6810
C	−1.4317	−39.064	−33.356
D	−17.7877	−0.2778	0.1279

Activation temperature: 2400 K, Ref. Temp.: 298 K, Pressure range: 0.010–100 kPa

Definitions of *B*, *C* and *D* here are a little different from those in the text

Although this derivation is stricter than the former, substitution of χ = 0.28 in Table 7.1 shows that the two tables are the same in practice. After these two derivations, the meanings of the thermodynamic data of superequilibrium atomic hydrogen and activated graphite can be understood more clearly. Thermodynamic data of superequilibrium atomic hydrogen must be calculated on the basis of the practical nonequilibrium concentration in its corresponding stationary states; that is the important key point of thermodynamic calculations of nonequilibrium systems.

It should also be noted that in both of these derivations the concentration of superequilibrium atomic hydrogen is defined as the equilibrium concentration of atomic hydrogen at the activated temperature. If the practical concentration of superequilibrium atomic hydrogen and the χ value can be experimentally measured, it may be found that the practical concentration of superequilibrium atomic hydrogen will be lower and the practical value will be higher. However, they will compensate each other, and will prevent too much deviation in the calculated results of (stationary) nonequilibrium phase diagrams.

7.5 Calculation Principle of Nonequilibrium Phase Diagrams

In classical thermodynamics phase diagrams were always regarded as equilibrium diagrams. There were only two basic kinds of phase diagrams, i.e. phase diagrams of stable equilibrium systems and phase diagrams of metastable equilibrium systems. Both of them are phase diagrams of equilibrium systems, and are called equilibrium phase diagrams for short. For instance, in a one-component p–T phase diagram a dashed line for supercooling liquid is sometime drawn below the triple point. This is a phase line of a metastable phase diagram. The essential prerequisite for such a dashed phase line is the nonexistence of the stable solid phase. Strictly speaking, when the metastable supercooling liquid phase line has been drawn, the gas-solid phase line and the solid-liquid phase line should no longer be drawn. However, for convenience of applications all of these lines are drawn together in the same phase diagrams. For clarity, stable phase lines are drawn by solid lines, while metastable phase lines are drawn by dashed lines. If the phase diagram is only a metastable phase diagram or only a stable phase diagram, then it is not necessary to do so.

Equilibrium phase diagrams correspond practically to systems without any macroscopic reaction or process, i.e. $[\mathrm{d_i}S_\mathrm{p} = \mathrm{d_i}S = 0]$, or to isothermal isobaric $[\left(\mathrm{d}G_\mathrm{p}\right)_{T,p} = (\mathrm{d}G)_{T,p} = 0]$, while nonequilibrium phase diagrams correspond to systems with $[\mathrm{d_i}S_1 < 0, \mathrm{d_i}S_2 > 0\ \&\ \mathrm{d_i}S = 0]$, or to isothermal isobaric systems with $[(\mathrm{d}G_1)_{T,p} > 0,\ (\mathrm{d}G_2)_{T,p} < 0 \&\ (\mathrm{d}G)_{T,p} = 0]$. These two kinds of phase diagrams belong to classical thermodynamics and modern thermodynamics, respectively. The corresponding calculation principles are Gibbs free energy minimization principle and nonequilibrium nondissipation principle, respectively. During system in its stable stationary state the illustration of relations between different phases is stationary nonequilibrium phase diagram, or stationary phase diagram, for short. Stationary phase diagram is the most useful kind of nonequilibrium phase diagram. There is no substantial difference in characters and calculation principles between calculations for nonequilibrium phase diagrams and for stationary nonequilibrium phase diagrams, so usually only the term of "nonequilibrium phase diagrams" is used. The activated low-pressure CVD diamond growth system is a typical system of stationary states, in which the growth rate of diamond crystals or films is rather slow (about 10 μm h^{-1} for hot filament process) and the CVD diamond growth system can remain steady for several hours, several days or even longer. It therefore provides an excellent practical example for research on stationary nonequilibrium phase diagrams. In the following discussions it can be seen that there are still two subclasses of stable phase diagrams and metastable phase diagrams in stationary nonequilibrium phase diagrams. It should be keep in mind that the difference between the equilibrium phase diagram and the (stationary) nonequilibrium phase diagram is distinct from the difference between the stable phase diagram and the metastable phase diagram. No confusion should be made between these two differences. In particular, we should be careful to distinguish a stationary nonequilibrium phase diagram from a metastable equilibrium phase diagram. In any system of stationary nonequilibrium phase diagram there is at least one macroscopic process, while in the system of metastable equilibrium phase diagrams there is no macroscopic process due to a lack of a stable phase or a hindrance for a chemical reaction.

Any stable phase diagram or metastable phase diagram calculated on the basis of Gibbs free energy minimization principle must be equilibrium phase diagram. The calculation of (stationary) nonequilibrium phase diagram must be on the basis of nonequilibrium nondissipation principle, i.e. $[\mathrm{d_i}S_1 < 0,\mathrm{d_i}S_2 > 0\&\mathrm{d_i}S = 0]$ or $[(\mathrm{d}G_1)_{T,p} > 0, (\mathrm{d}G_2)_{T,p} < 0\& (\mathrm{d}G)_{T,p} = 0]$ under isothermal isobaric conditions. Any practical complex system with macroscopic change must be dissipative system, i.e. $[\mathrm{d_i}S_1 < 0,\mathrm{d_i}S_2 > 0\&\mathrm{d_i}S > 0]$ or under isothermal isobaric conditions $[(\mathrm{d}G_1)_{T,p} > 0, (\mathrm{d}G_2)_{T,p} < 0\& (\mathrm{d}G)_{T,p} < 0]$. Nondissipative system, i.e. $[\mathrm{d_i}S_1 < 0,\mathrm{d_i}S_2 > 0\&\mathrm{d_i}S = 0]$ or under isothermal isobaric conditions $[(\mathrm{d}G_1)_{T,p} > 0, (\mathrm{d}G_2)_{T,p} < 0\& (\mathrm{d}G)_{T,p} = 0]$, is only an utmost situation or an idealized situation. Utmost situation or idealized situation is very important in thermodynamics. This is the boundary between possibility and impossibility, or the position of phase line in calculated phase diagrams. The practical calculations of nonequilibrium phase diagrams on the basis of nonequilibrium nondissipative thermodynamics will be illustrated in what follows.

7.6 Calculation Method of Nonequilibrium Phase Diagrams

The detailed calculation method for (stationary) nonequilibrium phase diagrams will be described and discussed from two different viewpoints as follows.

7.6.1 Detailed Calculation Steps

The whole discussion may be firstly divided into three steps

1. According to the reaction coupling theorem, a complete expression set of reactions, Eq. (7.1)–Eq. (7.4) or Eq. (**??**)–Eq. (**??**), for activated low-pressure diamond growth from the vapor phase can be written out. Based on experimental observations it can be shown that the existence of superequilibrium atomic hydrogen has no obvious influence on either diamond growth from the vapor phase or diamond etching in the vapor phase, while it does have significant influence on graphite etching. Obviously, this is due to graphite being a sp^2 unsaturated carbon structure which is very easy to react with superequilibrium hydrogen atoms and accelerates graphite etching. From the viewpoint of reaction coupling it may seem that due to the energy released by the reaction of the association of hydrogen atoms, Eq. (7.2) or Eq. (6.13), the chemical potential of graphite has been increased. Therefore, the whole reaction Eq. (6.14),$\mathrm{C(gra)} + \chi\ \mathrm{H}* = 0.5\chi\ \mathrm{H_2} + \mathrm{C(dia)}$ $(T_{\mathrm{activation}} \gg T, p \leqslant 10^5\ \mathrm{Pa})$, can be rewritten as reaction Eq. (7.3), C (gra*) = C (dia) $(T_{\mathrm{activation}} \gg T, p \leqslant 10^5\ \mathrm{Pa})$ and Eq. (7.4), C (gra*) = [C (gra) + $\chi\ (\mathrm{H}* - 0.5\ \mathrm{H_2})$]. In this way, the definition of activated graphite has been introduced.
2. As mentioned above, thermodynamic data of superequilibrium atomic hydrogen and activated graphite can be deduced.
3. Thermodynamic data of activated graphite are used instead of thermodynamic data of ordinary graphite, and then ordinary programs based on Gibbs free energy minimization principle can still be used for the calculation of stationary nonequilibrium phase diagrams. As the constant nonequilibrium concentration of superequilibrium atomic hydrogen has already been included in thermodynamic data of activated graphite, and as the concentration of superequilibrium atomic hydrogen will not be influenced by

programs based on Gibbs free energy minimization principle, the results are just in agreement with nonequilibrium nondissipation principle, $[\mathrm{d_i}S_1 < 0, \mathrm{d_i}S_2 > 0 \& \mathrm{d_i}S = 0]$ or $[(\mathrm{d}G_1)_{T,p} > 0,\ (\mathrm{d}G_2)_{T,p} < 0 \& (\mathrm{d}G)_{T,p} = 0]$ under isothermal isobaric conditions.

7.6.2 Different Kinds of Phase Lines

As there are many kinds of phase lines concerned with the low-pressure diamond growth from the vapor phase, the three calculation steps described above may still not be enough for an understanding in detail. The following discussion on calculation will be from another viewpoint of different kinds of phase lines.

7.6.2.1 Calculation of phase line between the gas phase and the stable solid phase in equilibrium phase diagrams

The calculation for the equilibrium phase line between the solid phase of graphite [C (gra)] and the gas phase in a low-pressure carbon-hydrogen system belongs to this kind of calculation. The basic reaction equation (or a group of equations) is

$$\mathrm{C(gra)} = \nu_1\mathrm{H_2} + \nu_2\mathrm{CH_4} + \nu_3\mathrm{C_2H_2} + \cdots, (\Delta G)_{T,p} = 0. \tag{7.18}$$

Note that molecular hydrogen is also one of the initial reactants, and ν_1 has a negative value. The reaction may be rewritten in a simpler form

$$\mathrm{C(gra)} = \{\nu_i\mathrm{C}_{X_i}\mathrm{H}_{Y_i}\}_{\mathrm{gra}}, (\Delta G)_{T,p} = 0. \tag{7.19}$$

Here, $\{\nu_i\mathrm{C}_{X_i}\mathrm{H}_{Y_i}\}_{\mathrm{gra}}$ represents hydrocarbons in the gas phase in equilibrium with the solid graphite phase. Molecular hydrogen ($X = 0$, $Y = 2$) and atomic hydrogen ($X = 0$, $Y = 1$) have been included. According to the mass conservation law, we get $\sum \nu_i X_i = 1$ and $\sum \nu_i Y_i = 0$. At constant temperature and pressure, $\Delta G = 0$ may represent a system in the applicability framework of classical thermodynamics being in equilibrium. The Gibbs free energy of the system will not decrease any more. Therefore, programs based on the Gibbs free energy minimization principle can be used for calculation of these stable phase lines in equilibrium phase diagrams.

7.6.2.2 Calculation of the phase line between the gas phase and the metastable solid phase in equilibrium phase diagrams

The calculation of the equilibrium phase line between the solid phase of diamond [C (dia)] and the gas phase in a low-pressure carbon-hydrogen system belongs to this kind of calculation. The basic reaction is

$$\mathrm{C(dia)} = \{\nu_i\mathrm{C}_{X_i}\mathrm{H}_{Y_i}\}_{\mathrm{dia}}, (\Delta G)_{T,p} = 0. \tag{7.20}$$

Here, $\{\nu_i\mathrm{C}_{X_i}\mathrm{H}_{Y_i}\}_{\mathrm{dia}}$ represents hydrocarbons in the gas phase in equilibrium with the solid diamond phase. Molecular hydrogen ($X = 0, Y = 2$) and atomic hydrogen ($X = 0, Y = 1$) have also been included. According to the mass conservation law, we get $\sum \nu_i X_i = 1$ and

$\sum \nu_i Y_i = 0$. For metastable equilibrium phase diagrams ordinary programs based on the Gibbs free energy minimization principle can still be used for the calculation. Note that the concentrations of hydrocarbons with respect to the diamond gas-solid phase line are different from that with respect to the graphite gas-solid phase line, i.e. $\{\nu_i \mathrm{C}_{X_i}\mathrm{H}_{Y_i}\}_{\mathrm{dia}} \neq \{\nu_i \mathrm{C}_{X_i}\mathrm{H}_{Y_i}\}_{\mathrm{gra}}$. It should also be pointed out that during the calculation of the metastable gas-solid phase line the stable solid phase must be excluded.

7.6.2.3 Calculation of both the stable gas-solid phase line and the metastable gas-solid phase line in (stationary) nonequilibrium phase diagrams

From the above discussion, it can be seen that during the calculation of the metastable gas-solid phase line the stable solid phase must be excluded. Otherwise, the gas-solid phase line for the metastable solid phase will not be obtained. This problem will not be met in the calculation of the gas-solid phase line for the stable solid phase. However, if the irrelevant metastable solid phase is still excluded in the calculation of the gas-solid phase line for the stable solid phase, there will be no influence on the calculated results. Due to the complexity of (stationary) nonequilibrium phase diagrams, the stable solid phase and the metastable solid phase may be changed in different temperature ranges. Therefore, it will be better to prevent interference during the calculation of any gas-solid phase line by ensuring that any irrelevant solid phase is previously excluded. In practical calculations of (stationary) nonequilibrium phase diagrams the following two gas-solid phase lines should be calculated, that is,

$$\mathrm{C(gra*)} = \{\nu_i \mathrm{C}_{X_i}\mathrm{H}_{Y_i}\}_{\mathrm{gra*}}, (\Delta G)_{T,p} = 0 \tag{7.21}$$

and

$$\mathrm{C(dia)} = \{\nu_i \mathrm{C}_{X_i}\mathrm{H}_{Y_i}\}_{\mathrm{dia}}, (\Delta G)_{T,p} = 0. \tag{7.22}$$

Here, $\{\nu_i \mathrm{C}_{X_i}\mathrm{H}_{Y_i}\}_{\mathrm{gra*}}$ represents hydrocarbons in the gas phase in nondissipative states, $(\Delta G)_{T,p} = 0$. If the possible stabilization effect of superequilibrium atomic hydrogen on diamond can be neglected, the diamond gas-solid phase line in (stationary) nonequilibrium phase diagrams will be the same as that in equilibrium phase diagrams. Therefore, the key point is about the calculation of the gas-solid phase line with respect to activated graphite. According to the previously deduced thermodynamic data of activated graphite, there will no difficulty in calculating the gas-solid phase line with respect to activated graphite. Because the graphite phase has already been activated, activated graphite in a rather larger range of temperature has become a metastable phase. After the calculation it can be concluded that the concentrations of hydrocarbons with respect to the diamond gas-solid phase line do not equal the concentrations of hydrocarbons with respect to the activated graphite gas-solid phase line, i.e. $\{\nu_i \mathrm{C}_{X_i}\mathrm{H}_{Y_i}\}_{\mathrm{dia}} \neq \{\nu_i \mathrm{C}_{X_i}\mathrm{H}_{Y_i}\}_{\mathrm{gra*}}$. However, it should be noted that the expression, $(\Delta G)_{T,p} = 0$ in Eq. (7.21) at constant temperature and pressure no longer means the system is in equilibrium, but only nonequilibrium nondissipation of the Gibbs free energy of the system. In the calculation procedure, the Gibbs free energy of the reaction system will not be able to decrease any further, so a routine program based on the minimization principle of the Gibbs free energy can still be used. In fact, the practical meaning of such a calculation has been changed into the nonequilibrium non-(Gibbs free energy)-dissipation principle at constant temperature and pressure, i.e. the nonequilibrium nondissipation principle. The superequilibrium atomic hydrogen data included in the data for activated graphite have already been fixed at a special nonequilibrium concentration, so the whole system is limited in a stationary state. That is, the corresponding phase diagram is a nonequilibrium (stationary) phase diagram, or a stationary phase diagram, for short.

In addition, as discussed above, usually the stable and metastable solid phases do not appear simultaneously in an equilibrium gas-solid system in an interesting temperature range. Therefore, the drawing of both stable and metastable gas-solid phase lines on the same equilibrium phase diagram is more convenient only for discussion. However, that is not the same for a system of stationary states for the diamond process. The stable solid phase and the metastable solid phase may appear individually or simultaneously in a gas-solid system of stationary states. For instance, in the activated low-pressure CVD diamond growth system, diamond growth and graphite etching may take place simultaneously, so the drawing of both stable and metastable gas-solid phase lines on the same phase diagram is not only allowed but also usually necessary. Moreover, the concentrations of gas species on the surfaces of stable and metastable solid phases and in the different parts of the gas phase may be different. All these characteristics are specific for the system in stationary states and for the (stationary) nonequilibrium phase diagram. Therefore, these characteristics of the calculated phase diagrams also confirm that these phase diagrams are not equilibrium phase diagrams but (stationary) nonequilibrium phase diagrams.

7.7 *T–X* Nonequilibrium Phase Diagrams for C–H System

To make a clear comparison, we first calculate a temperature-composition (*T–X*) equilibrium phase diagram of binary carbon-hydrogen (C–H) system on the basis of Gibbs free energy minimization principle, as shown in Fig. 7.7. In the equilibrium phase diagram there is a pure graphite phase region, or what can be called graphite deposition region (filled by a net-like pattern, in Fig. 7.7), between two phase lines. Inside the net-like patterned region, it is possible for graphite to be deposited or grown, but diamond growth is impossible. Graphite is a stable phase, while diamond is a metastable phase under low pressure, 4800 Pa (101.325 kPa = 1 atmosphere). This is a typical conclusion of equilibrium thermodynamics.

Now, thermodynamic data of activated graphite are used instead of the original standard thermodynamic data of graphite, but the software program on the basis of Gibbs free energy minimization principle is still used. A series of *T–X* nonequilibrium phase diagrams of binary carbon-hydrogen system are obtained, as shown in Fig. 7.8. Each of these nonequilibrium phase diagrams is related with extent of nonequilibrium. For the activated low-pressure diamond synthesis, the nonequilibrium condition is connected with the concentration of superequilibrium atomic hydrogen, and can be quantitatively represented by the coupling reaction rate ratio $\chi = r_2/r_1$. Figure 7.8 has another deeper meaning. In fact, Fig. 7.8 includes both equilibrium phase diagram ($\chi = 0$) and nonequilibrium phase diagrams ($\chi > 0$). It means that equilibrium phase diagrams may be regarded as a special example of nonequilibrium phase diagrams, or that nonequilibrium phase diagrams are an extension of equilibrium phase diagrams. That is, at least one more dimension has been increased for nonequilibrium phase diagrams in comparison with equilibrium phase diagram.

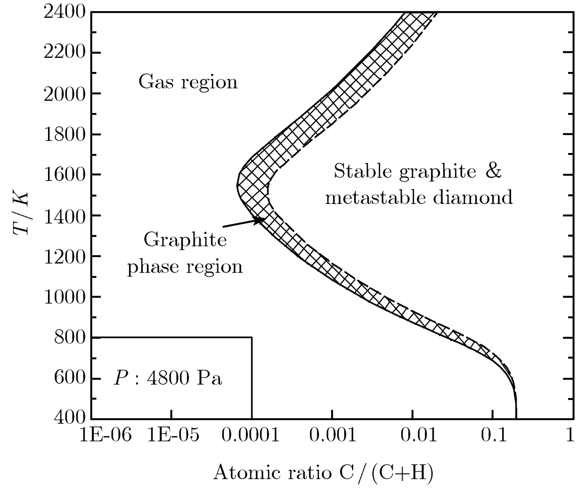


Fig. 7.7 Equilibrium phase diagram of carbon-hydrogen system

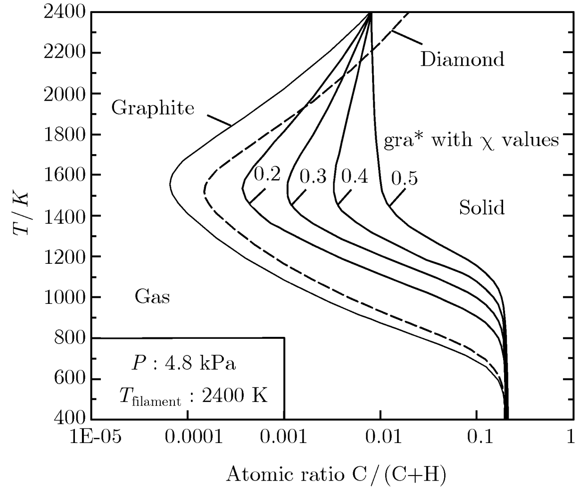


Fig. 7.8 Relationship between diamond growth region and reaction rate ratio χ in nonequilibrium phase diagrams

In calculation principles, equilibrium phase diagrams are practically based on $[\mathrm{d_i}S_\mathrm{p} = \mathrm{d_i}S = 0]$ or $[(\mathrm{d}G_\mathrm{p})_{T,p} = (\mathrm{d}G)_{T,p} = 0]$ for isothermal isobaric conditions, while nonequilibrium phase diagrams are based on $[\mathrm{d_i}S_1 < 0, \mathrm{d_i}S_2 > 0 \& \mathrm{d_i}S = 0]$ or $[(\mathrm{d}G_1)_{T,p} > 0, (\mathrm{d}G_2)_{T,p} < 0 \& (\mathrm{d}G)_{T,p} = 0]$ for isothermal isobaric conditions. For nonequilibrium phase diagrams, it can only be certain about $\mathrm{d_i}S_1 = -\mathrm{d_i}S_2$ or $(\mathrm{d}G_1)_{T,p} = -(\mathrm{d}G_2)_{T,p}$, and the extent of nonequilibrium or the absolute value of them is still not fixed. Therefore, $[\mathrm{d_i}S_\mathrm{p} = \mathrm{d_i}S = 0]$ or $[\left(\mathrm{d}G_\mathrm{p}\right)_{T,p} = (\mathrm{d}G)_{T,p} = 0]$ for isothermal isobaric conditions corresponds to only one equilibrium state, while $[\mathrm{d_i}S_1 < 0, \mathrm{d_i}S_2 > 0 \& \mathrm{d_i}S = 0]$ or $[(\mathrm{d}G_1)_{T,p} > 0, (\mathrm{d}G_2)_{T,p} < 0 \& (\mathrm{d}G)_{T,p} = 0$ for isothermal isobaric conditions may correspond to infinite nonequilibrium states. Therefore, the detailed experimental condition should be introduced to fix one of these nonequilibrium phase diagrams. That is the reason why one and only one experimental nonequilibrium parameter (such as χ here) should be introduced. In other words, for the activated low-pressure diamond growth, only one more dimension connected with concentration of superequilibrium atomic hydrogen in nonequilibrium phase diagrams has been increased than the dimensions in equilibrium phase diagram. In fact, from the definition of activated graphite C (gra*) = [C (gra) + χ (H*$-$ 0.5 H_2)], it can be found that χ value represents the extent of activation. The activation extent of graphite is an experimental nonequilibrium, but not a theoretical numerical value. A theoretical prediction of a minimum χ value is the threshold χ^e (χ^e is about 0.05, but related with temperature and pressure) to realize the transformation from graphite to diamond, but the χ value of a detailed experiment had better make a comparison between theoretical nonequilibrium phase diagrams and experimental results. That is the way accepted in this book.

Based on comparison between calculated curves with different χ values and experimental data reported by Matsumoto et al., [52] χ = 0.28 could be basically obtained, as shown in Fig. 7.9. It can also be found that the value of χ = 0.28 agrees well with diamond vapor growth experimental data reported by Sato [53] and Southworth [54]. All of their experimental data dropped just inside the diamond growth region of the theoretical calculated phase diagram. In the microwave process the temperature of neutral particles in the plasma is usually about 3000 K. That will be slightly favorable for increasing the diamond growth region. The thermodynamic data for activated graphite can be deduced again, but the χ value may still be 0.28. For the normal pressure plasma torch process the temperature of neutral particles in the plasma is much higher, so the concentration of superequilibrium atomic hydrogen close to the surface of the substrate may be higher. That is more favorable for increasing the χ value, and low-pressure diamond growth from the vapor phase. The plasma torch process runs at normal pressure. From (stationary) nonequilibrium phase diagrams at different pressures it can be found that the corresponding saturated concentration of carbon increases with the pressure. That is, the higher pressure is more favorable for increasing the growth rate from the vapor phase. The diamond growth rate at different places, even inside the same diamond growth region, is different. On the left-hand side close to the diamond phase line the supersaturation degree of carbon is lower. That is favorable for the growth of diamond single crystals, but unfavorable for the diamond nucleation. On the contrary, on the right-hand side close to the activated graphite phase line the super-saturation degree of carbon is higher. That is favorable for the diamond nucleation, but unfavorable for the growth of diamond single crystals. All of these theoretical predictions agree very well with the experimental observations.

Besides, Fig. 7.9 is also a (stationary) nonequilibrium phase diagram for a C–H system with the existence of superequilibrium of atomic hydrogen. It consists of a diamond gas-solid phase line (solid line in Fig. 7.9) and an activated graphite gas-solid phase line (dashed line in Fig. 7.9). The diagram has been divided into some regions by these two gas-solid phase lines and a horizontal line through intersection point (about 2050 K); that is (i) gas phase region (no deposition region, left side), (ii) diamond phase region (diamond growth region with oblique lines), (iii) a small graphite phase region at high temperature (black region), (iv) diamond stable and graphite metastable region (right side below intersection point) and (v) graphite stable and diamond metastable region (right side above intersection point). Both (iv) and (v) in diamond growth forum are usually called nondiamond carbon deposition regions.

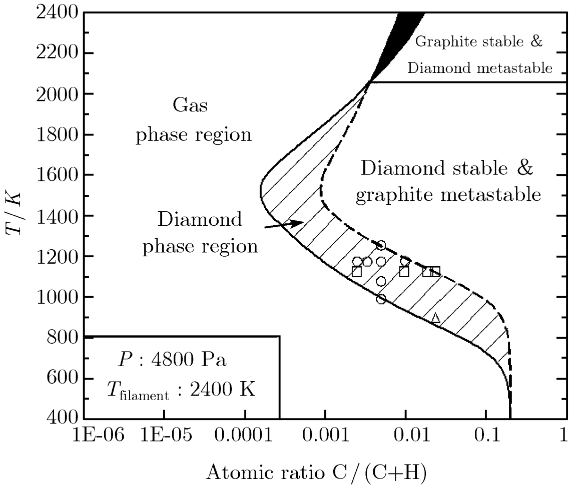


Fig. 7.9 *T–X* nonequilibrium phase diagrams of C-H system with the existence of superequilibrium atomic hydrogen. The symbols open circles, open squares and open triangles are experimental data reported by Matsumoto [52], Sato [53], and Southworth et al. [54], respectively

Note that above the intersection point of these two phase lines there is no diamond growth region. It means that if the substrate temperature is too high to close the hot filament temperature, due to the insufficient concentration of superequilibrium atomic hydrogen activated by the hot filament, the relative stabilities between diamond and graphite are not inverted. It also provides an explanation for the observation that if the hot filament temperature is too low to close the substrate temperature, the relative stabilities between diamond and graphite are not inverted. Just as reported in the literature, if the hot filament is made of platinum, diamond cannot grow in the hot filament process because of the low melting point of platinum (the melting point of pure platinum is 1773 °C, and that of platinum alloy is about 1650 °C). Figure 7.9 shows that below the temperature of the intersection point there is a diamond stable region (diamond growth region) between two gas–solid phase lines. Diamond has become a stable phase, while activated graphite has become a metastable phase. That is why diamond can grow from the vapor phase usually at a substrate temperature of 1000–1300 K, and why steady diamond growth together with simultaneous graphite etching can take place in the hot filament process.

Keep in mind that thermodynamic data of diamond, which can be found in databooks, are usually up to 1200 K. Thermodynamic data of diamond at higher temperature were obtained by extension without sufficient high accuracy in this book.

7.8 *T–p–X* Nonequilibrium Phase Diagrams for C–H and C–O Systems

It is well known that a complete phase diagram of C–H binary system should be expressed in a three-dimensional space of temperature-pressure-composition (*T–p–X*). Figure 7.10 is the calculated *T–p–X* phase diagrams of the C–H system. Three isobaric cross-sections of *T–X* phase diagrams at three different pressures (0.01 kPa, 1 kPa and 100 kPa) are shown in the figure. There is a diamond stable phase region (also called the diamond growth region) in each cross-section of the *T–X* phase diagram. The diamond growth phase region is represented by the region with oblique lines. If diamond growth is the purpose of our experiment, then the experiment should run inside the diamond growth phase region. That provides a powerful theoretical tool for designing and optimizing the activated low-pressure CVD diamond process. During the calculation, the empirical coupling reaction rate ratio $\chi = 0.28$ from the hot filament process has been used. The thermodynamic data for activated graphite for the C–H system has been listed in Table 7.1. This will not be repeated here.

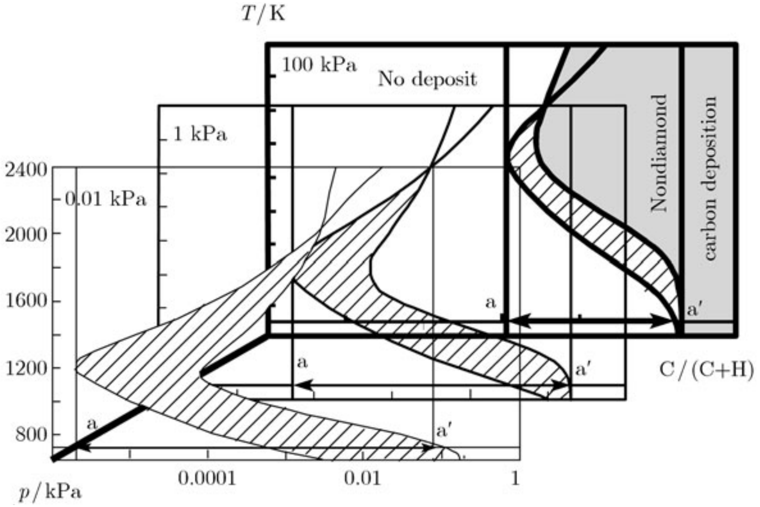


Fig. 7.10 *T-p-X* nonequilibrium phase diagrams of C-H system

Similarly, the research method for nonequilibrium phase diagrams of the C–H system can be extended for the C–O system, for which a set of coupled reaction equations can be written as

$$\mathrm{C(gra)} = \mathrm{C(dia)}, \Delta G_1 > 0 (T, p \leqslant 10^5\mathrm{Pa}). \tag{7.23}$$

$$\mathrm{O}* = 0.5\mathrm{O}_2, \Delta G_2 \ll 0 (T_{\mathrm{activation}} \gg T, p). \tag{7.24}$$

Here, O* represents atomic oxygen with superequilibrium concentration, or what can be called superequilibrium atomic oxygen (SAO). For the total reaction or for the whole system, an addition should be made between Eq. (7.23) and Eq. (7.24). Meanwhile, the reaction rate ratio of reaction Eq. (7.24) to reaction Eq. (7.23), $\chi_O = r_2/r_1$, should be under consideration. If χ_O is not too small, then [(7.25) = (7.23) + χ_O(7.24)] is as follows.

$$\mathrm{C(gra)} + \chi_{\mathrm{O}}\mathrm{O}* = 0.5\chi_{\mathrm{O}}\mathrm{O}_2 + \mathrm{C(dia)}, (T, p \leqslant 10^5\mathrm{Pa}) \tag{7.25}$$

or

$$\mathrm{C}\,[\mathrm{gra}*(\mathrm{O})] = \mathrm{C(dia)}, (T, p \leqslant 10^5\mathrm{Pa}). \tag{7.26}$$

Here,

$$\mathrm{C}\,[\mathrm{gra}*(\mathrm{O})] = [\mathrm{C(gra)} + \chi_{\mathrm{O}}(\mathrm{O}* - 0.5\mathrm{O}_2)]. \tag{7.27}$$

If the reaction rate ratio $\chi_O = r_2/r_1 >| \Delta G_1/\Delta G_2 |$ is not too small (i.e. atomic oxygen is of sufficient superequilibrium concentration), then we must have $\Delta G < 0$, and the reaction Eq. (7.25) or Eq. (7.26) should be toward the right side. gra*(O) represents graphite activated by superequilibrium atomic oxygen. The corresponding thermodynamic data of superequilibrium atomic oxygen and activated graphite (activated by superequilibrium atomic oxygen) for the C–O system can be deduced by a similar method used for the C–H system. The obtained thermodynamic data for activated graphite (activated by superequilibrium atomic oxygen) are listed in Table 7.6.

Table 7.6 Thermodynamic data for activated graphite (activated by superequilibrium atomic oxygen) [gra*(O)] with different χ_O values

$H_{\mathrm{gra}*}$(O) (kJ mol^{-1}) = $\chi_O \times 249.195$		
$S_{\mathrm{gra}*}$(O) (J mol^{-1} K^{-1}) = (5.6861 + $\chi_O \times 97.9819$)		
C_p (J mol^{-1} K^{-1}) $= A + B \times 10^{-3}T + C \times 10^5 T^{-2} + D \times 10^{-6} T^2$		
	298–1000 K	1000–1550 K
A	0.1597 + $\chi_O \times$ 8.4496	25.5701 + $\chi_O \times$ 8.4496
B	38.8890 − $\chi_O \times$ 7.3095	−0.2232 −$\chi_O \times$ 7.3095
C	−1.5154 + $\chi_O \times$ 0.7381	−39.1476 + $\chi_O \times$ 0.7381
D	−17.4121 + $\chi_O \times$ 2.2678	0.0978 + $\chi_O \times$ 2.2678
	1550–2000 K	2000–2400 K
A	25.5701 + $\chi_O \times$ 5.5648	25.5701 +$\chi_O \times$ 5.5439
B	−0.2232 − $\chi_O \times$ 2.1096	−0.2232 −$\chi_O \times$ 2.0594
C	−39.1476 − $\chi_O \times$ 5.4884	−39.1476 − $\chi_O \times$ 6.4633
D	0.0978 + $\chi_O \times$ 0.1755	0.0978 + $\chi_O \times$ 0.1755

Activation temperature: 2400 K, Ref. Temp.: 298 K, Pressure: 100 kPa.
Definitions of B, C and D here are a little different from those in text

T–X nonequilibrium phase diagrams of the C–O system at different pressures can then be calculated, and finally the calculated T–p–X nonequilibrium phase diagram of the C–O system is illustrated in Fig. 7.11. Although there are also three main regions (gas phase region, diamond growth region and nondiamond deposition region) in each T–X nonequilibrium phase diagram of the C–O system, they are quite different from the corresponding ones in T–p–X nonequilibrium phase diagram of the C–H system in both shape and position. The abscissa, i.e. the coordinate of the composition X, is selected by logarithms for the C–H system, but in the most general linear form for the C–O system. For the C–O system the temperature interval of the main part of the diamond growth region is obviously lower than the corresponding one for the C–H system. That agrees with a general empirical rule that diamond growth is usually at a somewhat lower temperature in an oxygen-containing system. The empirical rule can be illustrated very well by nonequilibrium phase diagrams for the C–O system.

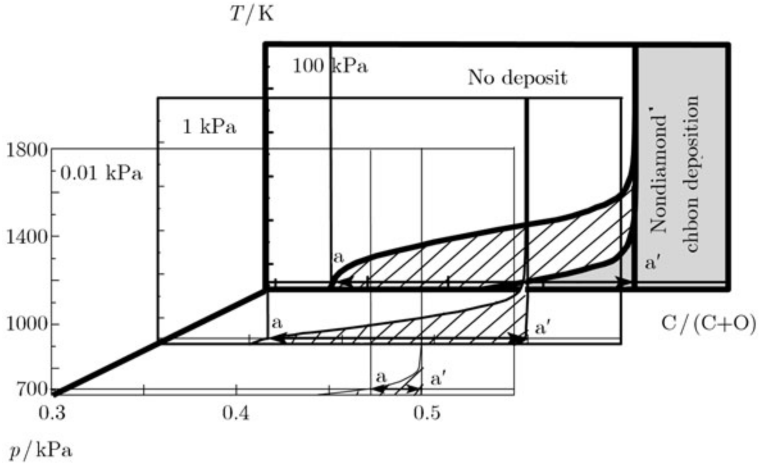


Fig. 7.10 *T*-*p*-*X* nonequilibrium phase diagrams of C-O system

It should be pointed out that because experimental data of the pure C–O system were very few, the reaction rate ratio χ_O used for calculation was simply from the approximation $\chi_O \approx \chi_H = 0.28$. Besides, another gas allotrope of oxygen, i.e. ozone, has been neglected to simplify the treatment.

7.9 *T*–*X* Nonequilibrium Phase Diagrams for C–(H+O) Systems

Just as for equilibrium phase diagrams, there are also different kinds of illustrations for nonequilibrium phase diagrams. For instance, nonequilibrium phase diagrams of C–H binary systems can also be illustrated as nonequilibrium phase diagrams of the CH_4-H_2 binary system. The differences between them are only on the meaning of the binary components and the scale of the coordinate axes. However, the form of phase diagrams of the CH_4-H_2 system may be more convenient for practical applications.[29, 31] In phase diagrams of the CH_4-H_2 binary system the composition of the most generally used 1 % CH_4—99 % H_2 gas system equals $X'=[CH_4/(CH_4 + H_2)] = 0.01$, but in phase diagrams of the C–H binary system the composition should be calculated by the transformation equation $X=[C/(C + H)] = [1/(1 + 4 + 2\times 99)] \approx 0.005$. Such a transformation is still very simple, so in this book phase diagrams of the CH_4-H_2 binary system will not be further discussed. The phase diagrams of CO–H_2, C_2H_6–CO_2, C_2H_4–CO_2, C_2H_2–CO_2, etc. binary systems with practical meanings [28,32,37~40] are also parts of the phase diagrams of the C–H–O ternary system, so they will not be further discussed here eithe

The following discussion will focus on nonequilibrium phase diagrams of a general pseudo-binary system with a fixed molar ratio [λ = O/(H+O)]. In fact, these systems are the C–H–O ternary system. A nonequilibrium phase diagram of C–H–O ternary system can be constructed on the basis of a series of *T*–*p*–*X* phase diagrams of a pseudo-binary system with different λ = 0.0, 0.1, 0.2, 0.3, ···, 0.9, and 1.0, respectively. That will be of benefit in the next chapter for construction of ternary nonequilibrium phase diagrams. To limit the length of the book, *T*–*p*–*X* phase diagrams of a pseudo binary system with

C–[O/(O+H)= λ] are illustrated only at three different pressures of 100 kPa, 1 kPa and 10 Pa, respectively in Fig. 7.12, Fig. 7.13 and Fig. 7.14. Moreover, a series of *T–X* phase diagrams of the pseudo binary system for λ = 0.0, 0.1, 0.2, 0.3, $\cdots$, 0.9 and 1.0, respectively, at the same pressure are drawn into one of these figures. It should be noted that if all the phase lines for λ = 0.0 are taken from Fig. 7.12, Fig. 7.13 and Fig. 7.14, respectively, then a three-dimensional *T–p–X* phase diagram of a C–H binary system similar to Fig. 7.10 can be constructed. The difference between the new constructed phase diagram and Fig. 7.10 is on the abscissa. In the new constructed phase diagram the abscissa of the composition is linear, while in Fig. 7.10 the abscissa of the composition is logarithm. In this way, a series of three-dimensional *T–p–X* phase diagrams of the C–(O+H) pseudo binary system with different λ values can be constructed. A three–dimensional *T–p–X* phase diagram of the C–(O+H) pseudo binary system with λ = 1

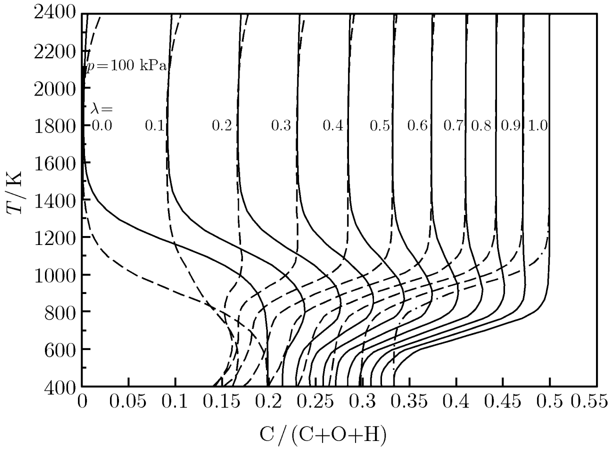


Fig. 7.12 *T-p-X* phase diagrams of the [C-(O+H)] pseudobinary system at 100 kPa

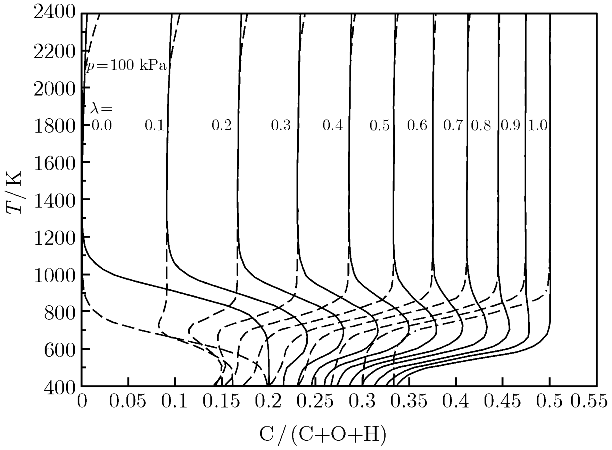


Fig. 7.13 *T-p-X* phase diagrams of the [C-(O+H)] pseudobinary system at 1 kPa

is just the same as the three-dimensional T–p–X phase diagram of the C–O pure binary system, i.e. Fig. 7.11. Details about how to construct these T–p–X nonequilibrium phase diagrams of C–H–O ternary systems will be discussed in the next chapter.

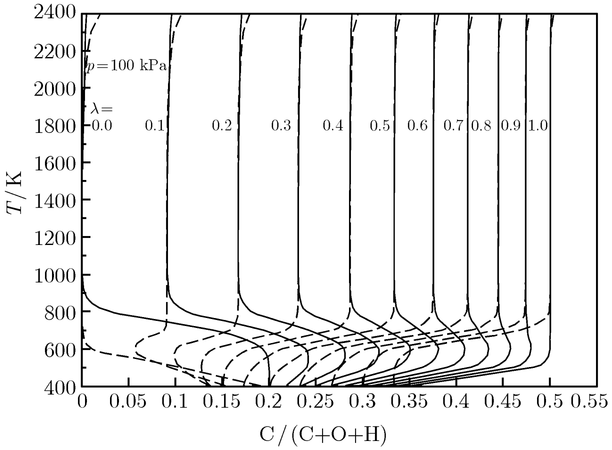


Fig. 7.14 T-p-X phase diagrams of the [C-(O+H)] pseudobinary system at 10 Pa

7.10 Gas Composition Nonequilibrium Phase Diagrams for C–H Systems

The discussions above focused on gas-solid phase diagrams. Gas-solid phase diagrams belong to one kind of phase diagram, which are mainly concerned with the saturation degree of each solid phase in the gas phase, such as the gas phase is undersaturated, saturated, or supersaturated to a solid phase. If the gas phase is undersaturated to the solid phase, the solid phase will be etched. If the gas phase is saturated to the solid phase, the solid phase will be in equilibrium or in stable stationary state. If the gas phase is supersaturated to the solid phase, the solid phase will be deposited or will grow. That is very useful to guide practical research. In these phase diagrams there is often a single gas-solid phase line to divide the whole diagram into a gas phase region and a solid phase region for each solid phase. However, such phase diagrams are usually not concerned with the detailed gas composition.

Another kind of phase diagram is mainly concerned with the gas phase composition. These are called gas phase composition T–p phase diagrams. Both kinds of gas–solid phase diagrams and gas phase composition T–p phase diagrams are complementary to each other. Figure 7.15 is an equilibrium gas phase composition phase diagram of a C–H system. The solid lines represent the saturation phase line of each gas species with respect to the graphite solid phase, while the dashed lines represent the saturation phase line of each gas species with respect to the diamond solid phase. In such phase diagrams there is always a saturation phase line for each gas species with respect to each solid phase. Therefore, there is a fixed value of saturated partial pressure for each gas species at different temperature and pressure for each solid phase. In this way the whole saturated gas phase composition is clearly known. In Fig. 7.15 all saturation phase lines of hydrocarbon gas species with respect to diamond (dashed lines) are higher than the corresponding saturation phase lines with respect to graphite. That is, diamond is a metastable phase, while graphite is a stable phase. That agrees with the conclusion of classical equilibrium thermodynamics from the gas-solid equilibrium phase diagrams of the C–H system at low pressure.

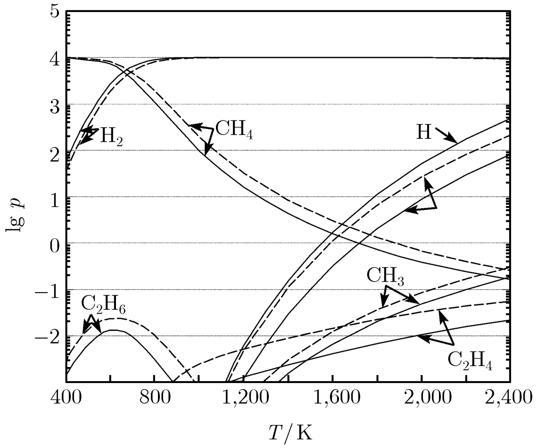


Fig. 7.15 Gas composition equilibrium phase diagrams for C-H system at p = 10 kPa (solid lines are phase lines of each gas species with respect to graphite, while dashed lines are phase lines of each gas species with respect to diamond)

Figure 7.16 is a gas composition T–p nonequilibrium phase diagram for activated low-pressure diamond growth from the vapor phase of the C–H system at the activated temper-

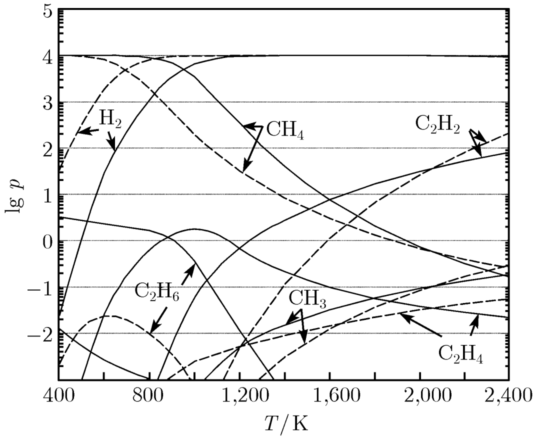


Fig. 7.16 Gas composition nonequilibrium (stationary) phase diagrams for activated low pressure diamond growth from a C-H system ($T_{\text{activation}}$ = 2400 K, p = 10 kPa) (solid lines are phase lines of each gas species with respect to activated graphite, while dashed lines are phase lines of each gas species with respect to diamond)

ature (such as the hot filament temperature) equal to 2400 K, and the total pressure equal to 10 kPa. In Fig. 7.16 at the substrate temperature of about 1200 K all saturation phase lines of hydrocarbon gas species with respect to diamond (dotted lines) are lower than the corresponding saturation phase lines with respect to activated graphite (solid lines). That is, diamond has become a stable phase under the activated low-pressure diamond growth condition, while graphite has been activated into a metastable phase. Therefore, both diamond growths from methane–hydrogen mixture, and from graphite via gaseous hydrocarbons under the activated low-pressure diamond growth conditions are thermodynamically reasonable. More detailed theoretical information about the saturated concentration (or the saturated partial pressure) of each gas species can be obtained from gas composition nonequilibrium phase diagrams. This is very favorable for knowing the main gas species, the growth mechanism and so on. Influences of gas composition on orientation of crystal growth will be discussed in the next section.

7.11 Influences of Gas Composition on Orientation of Crystal Growth

It is well known in the CVD diamond forum that different growth mechanisms of diamond crystals come from different gas precursors. CH_3 and C_2H_2 are the dominant growth precursors during activated CVD diamond growth. The growths of diamond (111) and (100) facets are controlled by C_2H_2 and CH_3, respectively. The ratio will affect the facets of diamond crystals. [55] As a general rule, in the case of the growth of separate crystals, the facets with the maximal growth rate disappear and the crystal habit represents the crystallographic facets with the lowest growth rate (Wulff's rule). When the ratio C_2H_2 to CH_3 (i.e. C_2H_2/CH_3]) exceeds a certain value, diamond (100) facets, whose growth is controlled by CH_3, will appear. Figure 7.17 shows the relationship between the concentration of C_2H_2 (or CH_3) and the substrate temperature on the basis of the nonequilibrium thermodynamic coupling model. The concentrations of C_2H_2 and CH_3 increase when the substrate temperature rises. The concentration of C_2H_2 rises more quickly.

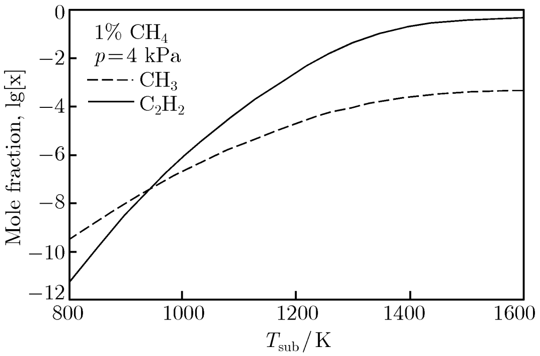


Fig. 7.17 The concentrations of CH_3 and C_2H_2 as the functions of substrate temperature during activated low-pressure diamond growth from the vapor phase

The relationship between the concentration ratio of [C_2H_2]/[CH_3] and the substrate temperature has also been calculated, as shown in Fig. 7.18. The curve represents 1 % methane in the feeding gas. In the normal range of the deposition temperature (900–1400 K), the ratio of [C_2H_2]/[CH_3] rises with increase of substrate temperature. Therefore, the growth ratio of diamond (100) to (111) facets (symbolized as $\alpha = r_{100}/r_{111}$) will fall with increase of the substrate temperature. Based on Wulff's theory, diamond (111) facets will appear when $\alpha \geqslant \sqrt{3}$, and diamond with (100) facets will be formed when $\alpha \leqslant \frac{1}{\sqrt{3}}$. It is concluded that the morphology of CVD separate diamond crystals shifts from initial (111) facets to (100) facets with the substrate temperature increasing, which is due to decrease of α with increase of [C_2H_2]/[CH_3] value. Mastumoto's experiments[52] had really showed that diamond crystal facets changed from (111) to (100) with the increase of the substrate temperature from 610 to 970 °C. That is, the theoretical calculations are well consistent with experimental observations.

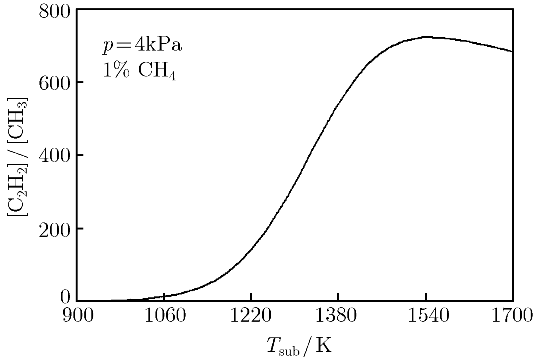


Fig. 7.18 The ratio of C_2H_2/CH_3 as a function of the substrate temperature during activated low-pressure diamond growth from the vapor phase

Figure 7.19 shows the dependence of [C_2H_2]/[CH_3] on the methane concentration at the same substrate temperature. The ratio [C_2H_2]/[CH_3] almost shows a linear increase with increase of the methane concentration. The ratio [C_2H_2]/[CH_3] has an opposite trend with the change of α. So when α decreases to a value smaller than $\frac{1}{\sqrt{3}}$ due to the ratio of

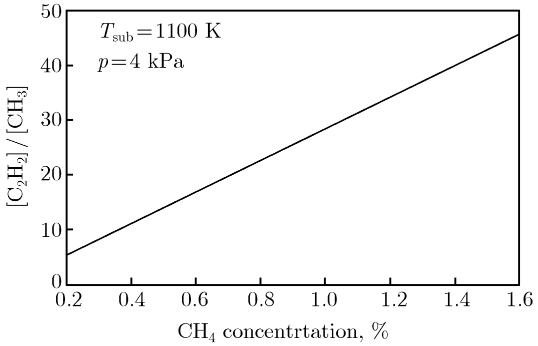


Fig. 7.19 The value of C_2H_2/CH_3 as a function of the methane concentration during activated low-pressure diamond growth from the vapor phase

$[C_2H_2]/[CH_3]$ increase with the CH_4 concentration, the morphology of diamond crystals will consequently be changed from (111) to (100). Kobashi et al.[56] used the different methane concentrations under 4 kPa pressure and 1100 K substrate temperature for diamond crystal growth. They found that when the methane concentration changed from 0.2 % to 1.2 % the morphology was initially (111) facet and subsequently (100) facet when the methane concentration was more than 0.4 %. Their experimental results indicate that our calculated results in Fig. 7.19 are reasonable. In addition, Sato et al.[57] also reported that when the methane concentration changed from low to high, the morphology changed from (111) facets to (100) facets at 0.5 % methane concentration.

However, in a polycrystalline diamond film, the growth of a community of microcrystals proceeds as a competition between separate crystals for feeding from the gas phase. In such a competition, prevalence and opportunity to be presented at the outer surface characterize the crystallites which have the maximal growth rate. Spitsyn has summarized experimental observations of orientation or texture of diamond films, as shown in Fig. 7.20.[58]

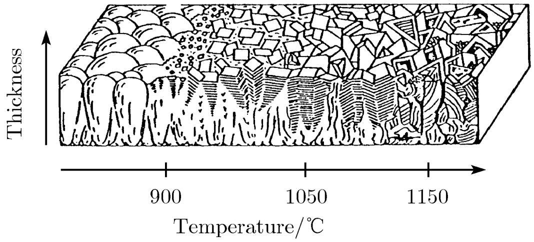


Fig. 7.20 Evolution of growth texture with diamond layer thickness and growth temperature. Reprinted from B.V. Spitsyn: 'Chap. 10 Growth of Diamond Films from the Vapor Phase'. In: *Handbook of Crystal Growth, vol.1*, ed. by D.T.J. Hurle (Elsevier Science, Amsterdam, 1994) pp. 403–456 by permission of Elsevier Science

At the low crystallization temperature or at the early stages of the growth of diamond films, nuclei consist of tiny polycrystalline diamond crystals, usually called "cauliflower-like" particles, as shown in the left of Fig. 7.20. With increase of the substrate temperature the situation may change drastically, as shown in Fig. 7.20. A $\langle 100\rangle$ orientated diamond film and a $\langle 111\rangle$ orientated diamond film may be obtained. $\langle 100\rangle$ orientated diamond films should be grown at a little lower temperature (about 950 °C with $\alpha = r_{100}/r_{111} > 1$), while $\langle 111\rangle$ orientated diamond films should be grown at a little higher temperature (about 1150 °C with $\alpha = r_{100}/r_{111} < 1$). These phenomena agree well with our theoretical calculation on the ratio of $[C_2H_2]/[CH_3]$, which has an opposite trend with the change of . In such a way, a lot of experimental observations on the morphology of diamond crystals and the orientation of diamond films in the literature can theoretically be summarized.

References

1. Wang J-T, Zheng P-J. The Second Law in Modern Thermodynamics. 19th Internat Conf on Chem Thermodynamics [C]. Boulder, CO, USA: 30 Jul - 4 Aug 2006.

http://www.symp16.nist.gov/pdf/p2086.pdf. Cited 20 Apr 2008.

2. Wang J-T. Modern thermodynamics – A whole view of thermodynamics [M]. Shanghai: Fudan Univ. Press, 2005.
3. Wang J-T. Modern thermodynamics – A whole view of thermodynamics. 2005 National Symp on Thermodynamics and Statistics [C], Dalian, China: 8 Aug 2005.
4. Wang J-T. A New Field of Nonequilibrium Nondissipative Thermodynamics (plenary lecture). Abstracts of Russian International Conf on Chemical Thermodynamics (RCCT2005) [C]. Moscow, Russia: 31 Jun - 2 Jul, 2005.
5. Wang J-T. New Progress in Studying the Second Law of Thermodynamics. Proc. of 12th National Symp. on Phase Diagrams [C], Shenzhen, China: 24-26 Oct 2004, Chinese Physical Soc, 274-278
6. Wang J-T. New Progress in Studying the Second Law of Thermodynamics. The 18th IUPAC Internat Conf on Chemical Thermodynamics (ICCT 2004) [C], Beijing: 17-21 Aug 2004.
7. Wang J-T, Zhang D W, Yu W-F. Progress in Studying the Second Law of Thermodynamics. Internat Conf on Phys Edu & Frontier Research (OCPA 2004) [C], Shanghai: 28 Jun - 1 Jul 2004.
8. Wang J-T, Zhang D W, Yu W-F. The Second Law of Thermodynamics for Ad-vanced Thin Films. Fifth Internat Conf on Thin Film Physics and Applications (TFPA2004) [C], “edited by Chu J-H, Lai Z-S, Wang L-W, Xu S-H”, Proc of SPIE [C], 2004, 5774: 23-28; (SPIE, Belingham, WA, 2004) Shanghai: 31 May - 2 Jun 2004.
9. Wang Z-X. Thermodynamics [M]. Beijing: High Education Press, 1955, 152-158.
10. Su R-K. Statistics [M]. 2nd Ed. Beijing: High Education Press, 2004, 257
11. Wang J-T. The Second Law of Thermodynamics in the Current 21st Century (plenary lecture). 2007 National Symp on Thermodynamics and Statistics [C], Yanji, China: 6 Aug 2007.
12. Wang J-T, Zhang D W. Nonequilibrium Nondissipative Thermodynamics & Calculation of Nonequilibrium Phase Diagrams. In: Proc. of 11th National Symp. on Phase Diagrams [C]. Xining: 20–24 Aug 2002, 131-134; J Salt Lake Research [J], 2003, 11(1): 62-65.
13. Wang J-T, Zhang D W. New Stage of Modern Thermodynamics Promoted by Low-pressure Diamond Syhthsis. 2002 Proc of Chinese MRS (abstract) [C], CMRS, Beijing: 21-24 Oct 2002, E56.
14. Wang J-T. Physics (Beijing) [J], 2003, 32(1): 9-15.
15. Wang J-T. Fundamental Progress in Thermodynamics. 2003 National Symp on Thermodynamics and Statistics [C], Huhehaote, China: 18 Aug 2003.
16. Wang J-T. Nonequilibrium Nondissipative Thermodynamics – With Application to Low-Pressure Diamond Synthesis [M]. Heidelberg: Springer, 2002.
17. Wang J-T. Phase Diagrams of Stationary Nonequilibrium States — Thermodynamics for Low Pressure Diamond Growth from the Vapor Phase [M]. Beijing: Science Press, 2000, 212 pages.
18. Wang J-T, Zhang D W, Liu Z-J. Thermodynamic Coupling Model for Low Pressure Diamond Growth from the Vapor Phase [M]. Beijing: Science Press, 1998, reprinted in 2000, 175 pages.
19. Wang J-T, Wang H, Q-H. Yang, Zheng P-J. ’94 Fall Proc of Chinese MRS (III) [C], Beijing: Chem Ind Press, 1995, 190-195.
20. Wang J-T. Proc of ’94 Fall Chinese MRS (IV) [C], New Technology for Materials Process and Research (II), Beijing: Chem Ind Press, 1995, 496-499.
21. Wang J-T, Zheng P-J. Chinese Science Bulletin [J], 1995, 40(11): 1056; Eng ed. 1995, 40(13): 1141-1143.
22. Wang J-T, Zheng P-J, Yang Q-H, Wang H. Phase diagram calculations of nonequilibrium stationary states for CVD diamond growth. In: Diamond Materials, 4th Internat Symp [C], ed by Ravi K V & Dismukes J P, The Electrochem Soc, 1995, PV95-4: 13-19.

23. Wang J-T. Proc of 13th International Conf on CVD [C], The Electrochem Soc Inc, Los Angeles: May 1996, PV96-5: 651-655.
24. Wang J-T, Huang Z-Q, Yang Q-H, et al. Proc of 13th International Conf on CVD [C], The Electrochem Soc Inc, Los Angeles: May 1996, PV96-5: 727-732.
25. Wang J-T, Zhang D W, Wan Y-Z, Huang Z-Q. Presented as paper no. 8.092b at Diamond '96 Conf [C], Tours, France: 8-13 Sept 1996.
26. Y-Z. Wan, Zhang D W, Wang J-T, Huang Z-Q. Presented at '96 CMRS Meeting, Section C Beijing; '96 Proc of Chinese MRS, Functional Materials (I-3) [C], Beijing: Chemical Industry Press, 1997, 552-555
27. Wang J-T, Wan Y-Z, Zhang D W, Liu Z-J, Z-Q. Huang, J Materials Research [J], 1997, 12(12): 3250-3253.
28. Zhang D W, Wan Y-Z, Wang J-T. J Crystal Growth [J], 1997, 177(1/2): 171-173.
29. Zhang D W, Wan Y-Z, Liu Z-J, Wang J-T. J Mater Sci Lett [J], 1997, 16: 1349-1351.
30. Wang J-T, Wan Y-Z, Zhang D W, Huang Z-Q, Zheng P-J. Prog in Natural Science [J], 1997, 7(4): 392-396; Eng Ed. 1997, 7(3): 265-271.
31. Zhang D W, Wan Y-Z, Wang J-T. Prog in Natural Science [J], 1997, 7(2): 170-174; Eng Ed. 1997, 7(4): 478-482.
32. Zhang D W, Wan Y-Z, Wang J-T. Prog in Natural Science [J], 1997, 7(5): 634-637; Eng Ed. 1997, 7(4): 506-508.
33. Wan Y-Z, Zhang D W, Liu Z-J, Wang J-T. High Technol Lett [J], 1997, 3(2): 80-83.
34. Zhang D W, Wang J-T, Wan Y-Z, Acta Physica Sinica (Chinese Physics) [J], 1997, 46(6): 1237-1242.
35. Zhang D W, Wan Y-Z, Liu Z-J, Wang J-T. Acta Metallurgica Sinica [J], 1997, 33(11): 1189-1193.
36. Zhang D W, Wan Y-Z, Wang J-T. Chinese J of Inorg Mater [J], 1997, 12(3): 331-335.
37. Zhang D W, Wan Y-Z, Wang J-T. Chinese J of Inorg Mater [J], 1997, 12(4): 617-619.
38. Wang J-T, Wan Y-Z, Liu Z-J, Zhang D W. Proc of 9th National Symp on Phase Diagrams [C], Beijing, 1997, 17-19.
39. Wang J-T, Wan Y-Z, Liu Z-J, H Wang, Zhang D W, Huang Z-Q. Mater Lett [J], 1998, 33(4): 311-314.
40. Zhang D W, Liu Z-J, Wan Y-Z, Wang J-T. Appl Phys A. [J], 1998, 66(1): 49-51.
41. Wan Y-Z, Zhang D W, Liu Z-J, Wang J-T. Appl Phys A. [J], 1998, 67(2): 225-231.
42. Wan Y-Z, Zhang D W, Liu Z-J, Zhang J-Y, Wang J-T. Mater Chem Phys [J], 1998, 56(3): 275-279.
43. Wan Y-Z, Zhang D W, Liu Z-J, Wang J-T. J Synthetic Crystals [J], 1998, 27(2): 109-113.
44. Liu Z-J, Zhang D W, Wan Y-Z, Wang J-T. J Inorg Mater [J], 1998, 13(3): 432-434.
45. Liu Z-J, Zhang D W, Wan Y-Z, Wang J-T. J of Inorg Chem [J], 1998, 14(4): 412-417.
46. Liu Z-J, Wan Y-Z, Zhang D W, Wang J-T. Chemical J of Chinese Univ [J], 1998, 19(7): 1140-1143.
47. Liu Z-J, Zhang D W, Wan Y-Z, Wang J-T. Microfabrication Tech [J], 1998, 1998(1): 69-73.
48. Liu Z-J, Zhang D W, Wan Y-Z, Wang J-T. Functional Materials [J], 1998, 29(5): 506-508.
49. Wang J-T, Zhang D W, Liu Z-J, Wan Y-Z, Zhang J-Y. 27th Internat. Conf. on Phase Diagrams and its Application [C], Beijing: 1998.
50. Wang J-T. Physics (Beijing) [J], 1998, 27(2): 77-83.
51. Stull D R, et al. JANAF Thermochemical Table [DB]. 2nd ed. Washington: US Government Printing Office, 1971.
52. Matsumoto S, Sato Y, Tsutsumi M, Setaka N. J Mater Sci [J], 1982, 17: 3106-3112.
53. Sato Y, et al. Surface and Coatings Tech [J], 1989, 39/40: 183-198.
54. Southworth P, et al. In: Diamond Materials. 3rd Internt. Symp. at Honolulu, Hawaii [C], 1993, 87-94.

55. Zhang J-Y, Wang P-F, Ding S-J, Zhang D W, Wang J-T, Liu Z-J. Thin Solid Films [J], 2000, 368(2): 266-268.
56. K. Kobashi, K. Nishimura, Y. Kawate, et al. Phys. Rev. B [J], 1988, 38: 4067.
57. Y. Sato, M. Kamo, Surf Coat Technol [J], 1998, 39/40: 183.
58. Spitsyn B V. Chap. 10 Growth of Diamond Films from the Vapor Phase. In: Handbook of Crystal Growth [M]. ed by Hurle D T J. Amsterdam, NL: Elsevier Science, 1994, vol 1: 441.

Chapter 8

Nondissipative Thermodynamics and Ternary Nonequilibrium Phase Diagrams

Abstract Since 1986, a great upsurge in research on activated low-pressure diamond growth from the vapor phase has unfolded throughout the world. The gas sources for low-pressure diamond growth were also extended from C–H binary systems into C–H–O ternary systems. Bachmann's famous empirical phase diagram (1991) can be explained and improved by our quantitatively calculated projective nonequilibrium phase diagrams, while Marinelli's critical empirical phase diagram (1994) was near to quantitative agreement with our cross-section nonequilibrium phase diagram on the basis of thermodynamic coupling model. Moreover, theoretical C–H–X ternary nonequilibrium phase diagrams (X = F or Cl) are also in agreement with experiments reported in the literature. Due to the similarities between allotropic pairs of diamond–graphite and cubic boron nitride–hexagonal boron nitride (cBN–hBN), the new kind of nonequilibrium phase diagrams can also be used for activated CVD cBN processes. Some nice evaluations of nonequilibrium phase diagrams can also be found in the literature.

8.1 Bachmann's Empirical Phase Diagram

Since 1986, a great upsurge in research on activated low-pressure diamond growth from the vapor phase has unfolded throughout the world. Many scientists coming from different academic fields joined in this research and research papers had increased very quickly. At the early stage, the gas sources of carbon were mainly CH_4–H_2 mixture, and other hydrocarbons in a hydrogen mixture. The general expression could be written as C_xH_y–H_2. A little later, alcohol–H_2, methanol–H_2, acetone–H_2, ether–H_2, and other oxygen-containing organic compounds in hydrogen mixture were also used as reaction sources. The CH_4–H_2 mixture together with direct addition of a little oxygen was also used for diamond growth. It was observed that the addition of oxygen in the reaction sources is often favorable to get diamond thin films with higher quality and to lower the temperature for diamond growth. Sometimes, colorless transparent diamond films could be obtained. In the oxygen–acetylene torch combustion process for diamond thin film growth, oxygen plays a more important role. From thermodynamics point of view all of them belonged to C–H–O ternary systems.

In 1991, P.K. Bachmann et al. summarized reliable experimental data for activated CVD diamond growth for C–H–O systems reported in the literature.[1, 2] The summary included all kinds of different processes, such as the hot filament process, combustion process, microwave process, DC or AC plasma process, plasma torch process, and so on, without distinction. In this summary the experimental data (i.e. the C, H and O compositions of the gas sources) were divided into three classes: no growth (or no deposition), diamond growth, and nondiamond carbon deposition. The experimental points, where nothing at all was deposited, were only labeled by the number, *n*, of data points; the experimental points of diamond growth were marked by open diamond symbols; and the experimental points, where nondiamond carbon was deposited, were marked by full circles. The system composition (i.e. carbon, hydrogen, oxygen concentrations) can be represented as usual in a planar triangular phase diagram. It was found that the three classes of experimental points were concentrated into three corresponding regions inside the triangular ternary phase diagrams. These regions were called the no growth region, the diamond growth region (Bachmann called it the "diamond domain"), and the nondiamond carbon growth region. In this way, an atomic compositional C–H–O diamond deposition phase diagram with "diamond domain" was obtained, as shown in Fig. 8.1. In fact, such a phase diagram was an empirical phase diagram for activated low-pressure diamond from the vapor phase of the C–H–O system and was comprised of more than 80 deposition experiments from over 25 references. Obviously, the system composition position are independent of the initial raw materials inside the phase diagram. For instance, the position of CH_3OH is located at the same position as the 33.3 % CO and 66.6 % H_2 mixture, and also at the same position as the 66.6 % CH_4 and 33.3 % O_2 mixture. Moreover, the composition posi-

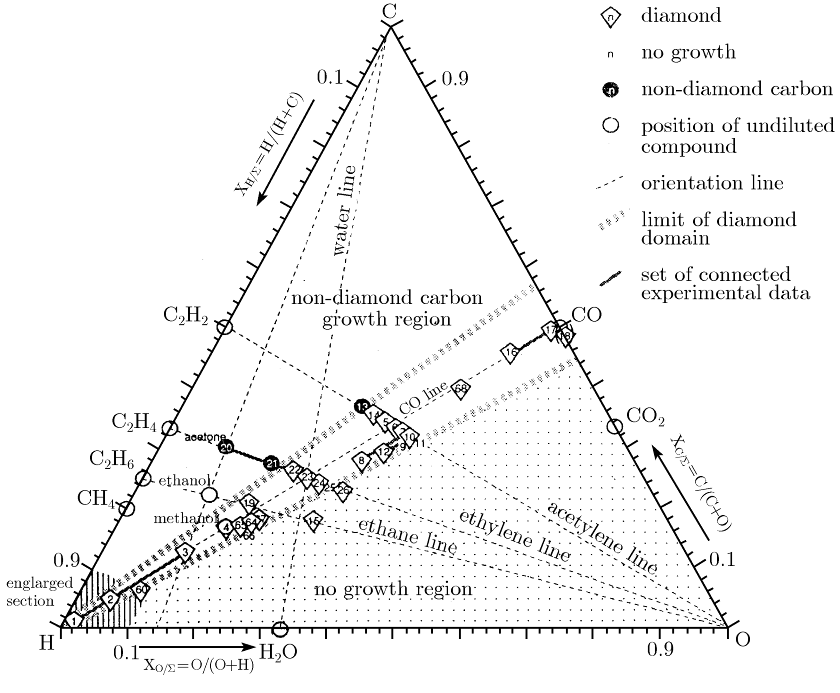


Fig. 8.1 Bachmann's empirical phase diagram for diamond deposition from C–H–O systems. Reprinted from P.K. Bachmann, D. Leers and H. Ldtin (Figure 1, p. 4) [1] by permission of Elsevier Science

tion inside the phase diagram for an experimental point is also independent of the process used. That is, all different processes, such as the hot filament process, microwave process, combustion process, and plasma torch process, were treated in the same way.

This characteristic that the phase diagram depends only on the total composition of the system but not on the initial raw materials and the processes used clearly showed that activated low-pressure diamond growth must be subordinated to thermodynamic rules. The phase diagram itself is a concept and a term of thermodynamics. There was a viewpoint for a long time that the activated low-pressure diamond growth from the vapor phase was controlled only by kinetics but not by thermodynamics. The reason may be based on a conclusion of equilibrium thermodynamics that "diamond is a metastable phase at low pressure". Now the existence of the diamond growth phase region clearly shows that such a viewpoint is incorrect for the activated low-pressure diamond process. The existence of the diamond growth phase region has also forced some scientists who incorrectly took the kinetic control viewpoint to consider the thermodynamics aspects. That was the advantage of Bachmann's empirical phase diagram.

However, the most important disadvantage was that the influences of temperature and pressure were not included in Bachmann's empirical phase diagram. From the division of phase regions, it can be clearly found that Bachmann had made an effort to exclude experimental points of no deposit and nondiamond carbon deposition out of the diamond growth phase region, i.e. Bachmann's "diamond domain". Because the influences of temperature and pressure were not considered, the shape of Bachmann's diamond growth region must be incorrect. Further discussions on it will be made in the following sections.

8.2 Projective Nonequilibrium Phase Diagrams for C–H–O Systems

As mentioned above, the composition of a ternary system in a phase diagram should be expressed by two independent variables, i.e. the composition should be expressed in a planar phase diagram. The substrate temperature T and the total pressure p need two additional coordinates. Therefore, a four-dimensional T–p–X_1–X_2 phase diagram of a ternary system cannot be completely expressed in ordinary three-dimensional space. To express such a ternary system in a three-dimensional diagram it is usual to fix one of these variables, such as pressure or temperature, and then a phase diagram of T–X_1–X_2 or p–X_1–X_2 can be drawn in three-dimensional space. Even so, such three-dimensional phase diagrams are not very convenient for quantitative applications. It is usual to introduce projective phase diagrams or cross-section phase diagrams. The common advantage of these two kinds of planar phase diagrams is that it is easy to compare between theory and experiments, and to offer a theoretical guide for practical work. However, these two different planar phase diagrams have their own advantages and disadvantages. For example, the projective planar diagram can include a lot of information for an overview. Projective planar diagrams may still be divided into three subclasses: (i) projective diagrams along the temperature coordinate within a certain temperature range at a fixed pressure; (ii) projective diagrams along the pressure coordinate within a certain pressure range at a fixed temperature; and (iii) projective diagrams both along the temperature coordinate within a certain temperature range and along the pressure coordinate within a certain pressure range. Among theses diagrams the third subclass of projective planar phase diagrams includes the largest amount of information. Figure 8.2 is such a theoretical projective nonequilibrium phase diagram of a C–H–O system calculated by us on the basis of the nonequilibrium nondissipation principle within the temperature range 700–1900 K and within the pressure range 0.01–100 kPa. [3~36] Ranges are selected so that most of the experimental data for activated low-pressure diamond growth from the vapor phase reported in the literature can be included. The selected ranges are also close to the ranges selected by Bachmann (>400 °C, 0.1–760 Torr), so that a comparison between our theoretically calculated phase diagram and Bachmann's empirical phase diagram can be made. Therefore, Bachmann's empirical phase lines (dotted lines) are also drawn in our theoretically calculated phase diagrams, as shown in Fig. 8.2. [25~27,32]

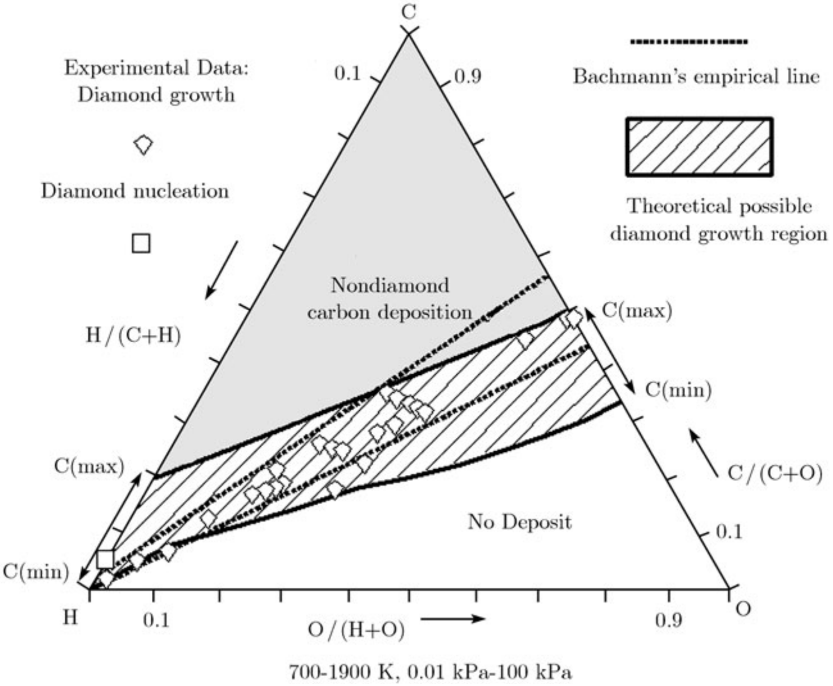


Fig. 8.2 Theoretically calculated projective nonequilibrium phase diagram of C–H–O ternary systems

There are also three phase regions in the theoretically calculated projective phase diagram of the C–H–O ternary system: (i) gas region (also called no deposition region); (ii) possible diamond growth region (note that in projective phase diagrams the name "possible diamond growth region" should be used. If the temperature and/or pressure are not suitable, then diamond growth may not occur); (iii) nondiamond carbon deposition region. These phase regions are divided by the carbon concentration limitation phase lines. The low concentration limitation line is located between the no deposition region and the possible diamond growth region, while the high concentration limitation line is located between the possible diamond growth region and the nondiamond carbon deposition region. The theoretically calculated low carbon concentration limitation line is much lower than that of Bachmann's empirical line, so the theoretically possible diamond growth region includes more experimental data than that in Bachmann's empirical phase diagram. Therefore, all the relatively reliable experimental data are included in the possible diamond growth region, but the empirical phase diagram did not agree with them. For the high carbon concentration limitation the theoretical line obviously crosses the empirical line, so that there is a small triangular region at each of the right and the left sides. The right triangle is located below Bachmann's empirical phase line (inside Bachmann's diamond growth region) but above the theoretical phase line (outside our theoretical possible diamond growth region). That is, in the right-hand small triangular region, i.e. a molar ratio greater than C/O = 1, diamond growth is impossible based on our theoretical phase diagram, but diamond growth would be possible based on Bachmann's empirical phase diagram. However, up to now no diamond growth experiment has yet been reported there in the literature. On the contrary, the left triangle is located above Bachmann's empirical phase line (outside Bachmann's diamond growth region) but below the theoretical phase line (inside our theoretical possible diamond growth region). That is, in the left small triangular region diamond growth is possible based on the theoretical phase diagram, but diamond growth would be impossible based on Bachmann's empirical phase diagram. In recent years, experiments had confirmed that in the left small triangular region diamond growth is possible, and diamond nucleation is also possible on a nondiamond substrate, as shown in Fig. 8.2 by an open square close to H apex.[37] In this way the correctness of our theoretical phase diagram had been further confirmed. The author met Bachmann himself during an international conference, and discussed this with him. Bachmann said he had already found these problems, and he had shifted the high limitation phase line of the left part approach to the position C/O = 1 (as shown in Fig. 8.3 of this book). His related paper was published in the proceedings of the conference.[38, 39] Obviously, if all reliable experimental data selected by Bachmann in his original empirical phase diagrams (i.e. Fig. 8.1 of this book) were put in the revised phase diagram, then much of the diamond growth data would drop into the revised "no growth region". Moreover, there was no theoretical basis supporting them to do so.

There is an obvious disadvantage for all projective phase diagrams. That is, during the projection different phase regions in space may partially overlap each other on the projective plane. For instance, the boundary of the projective diamond growth phase region is selected as a maximum borderline. Therefore, all diamond growth experimental points should be dropped into the projective diamond growth composition phase region, but maybe some parts of the no deposition phase region (the gas region) and nondiamond carbon deposition region have also been projected into the projective diamond growth composition phase region during projection. That is, the temperature and/or the pressure may be unsuitable for diamond growth. Due to this uncertainty, the projective diamond growth phase region should be called the "possible diamond growth phase region". Therefore, a few experimental points with no deposition and a few experimental points of nondiamond carbon deposition drop inside the possible diamond growth region in Fig. 8.2. That is normal and reasonable.

In cross-section phase diagrams all the phase lines and phase boundaries are clear and fixed. This is different from projective phase diagrams. However, there are also other disadvantages of cross-section phase diagrams. For instance, for each practical experimental

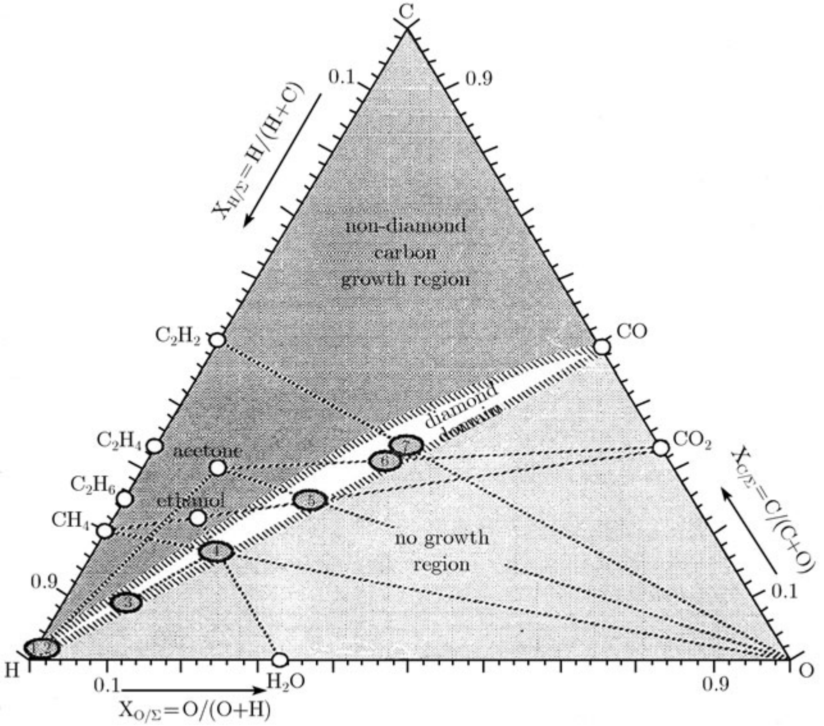


Fig. 8.3 Revised Bachmann's empirical phase diagram for diamond deposition from C–H–O systems. Reprinted from P.K. Bachmann, H.J. Hagemann, H. Lade et al. (Figure 1, p. 821) [39] by permission of Elsevier Science

condition a special cross-section phase diagram for the corresponding temperature and pressure should be drawn. That is not favorable for giving an overview of comparison between theoretical predictions and experiments. In this book the projective phase diagrams for C–H–O systems will be discussed first, and the discussion on cross-section phase diagrams will be made next. The calculation method is the same for both the projective phase diagram and the cross-section phase diagram, but the calculation range and treatment of data are different.

The following discussion on Fig. 8.2 will be taken as an example of projective phase diagrams for C–H–O systems. There are two sidelines of C–H and C–O in the C–H–O triangular composition phase diagram. The different phase ranges on these two sidelines (C–H and C–O) are obtained by two projections from T–p–X phase diagram of C–H and from T–p–X phase diagram of C–O systems, respectively. For instance, if Fig. 7.10 or Fig. 7.11 is projected along the T coordinate from 1900 K onto the plane of 700 K, then after the projection the diamond growth region (represented by regions with oblique lines) of each T–X phase diagram at different pressures will be a line segment on the composition X coordinate. These line segments are composition ranges for possible diamond growth (i.e. the possible diamond growth composition ranges), which are denoted as a–a' line segments at all the selected pressures in Fig. 7.10 (or Fig. 7.11). In the next step, both the minimum value of a in the selected pressure range (0.01–100 kPa), i.e. $a_{\min}$, and the maximum value of a' in the selected pressure range (0.01–100 kPa), i.e. a_{max}, are drawn on the corresponding sideline of C–H (or C–O) of the triangular phase diagram, as shown in Fig. 8.2, and denoted as C (min) and C (max), respectively.

For the calculation of the inner part of the C–H–O triangular phase diagram of the ternary system, a linear combination approximation is adopted for getting the thermodynamic data for graphite simultaneously activated by superequilibrium atomic hydrogen and superequilibrium atomic oxygen. For example, the thermodynamic data for graphite simultaneously activated by both superequilibrium atomic hydrogen and superequilibrium atomic oxygen with the atomic ratio H/O = 1 for the C–H–O system can be taken as the average of both the corresponding thermodynamic data for graphite activated by superequilibrium atomic hydrogen [gra*(H)] (see Table 7.1 and note that gra* $\equiv$ gra*(H), $\chi \equiv \chi_{H}$) and the corresponding thermodynamic data for graphite activated by superequilibrium atomic oxygen [gra*(O)] (see Table 7.6). Thermodynamic data for graphite simultaneously activated by superequilibrium atomic hydrogen and superequilibrium atomic oxygen with other H/O atomic ratios of the C–H–O system can be obtained by a linear insertion method. After getting the thermodynamic data for graphite simultaneously activated by superequilibrium atomic hydrogen and superequilibrium atomic oxygen with any fixed H/O atomic ratio, the *T–X* or *T–p–X* nonequilibrium phase diagrams of the C–(H/O) pseudo binary system (refer to Fig. 7.12 in Chapter 7) with any fixed H/O atomic ratio can be calculated. After two projections along the *T* coordinate and along the *p* coordinate, respectively, the line segment a_{min}–a'_{max} can be obtained. All these a_{min}– a'_{max} line segments with a fixed H/O ratio are next drawn into the planar triangular C–H–O composition phase diagram. The locus of all these a_{min} points is just the low carbon concentration limitation line C (min)–C (min) for possible diamond growth, while the locus of all these a'_{max} points is just the high carbon concentration limitation line C (max)–C (max) for possible diamond growth. In this way, Fig. 8.2 of the projective phase diagram of C–H–O ternary systems is obtained. According to the construction of the phase diagram mentioned above, it is clear that the planar phase diagram of Fig. 8.2 is a theoretical ternary projective composition phase diagram, while the corresponding Bachmann's phase diagram is an empirical ternary projective composition phase diagram. It is also clear that the projective composition nonequilibrium phase diagram of a ternary system, as shown in Fig. 8.2, was calculated on the basis of the nonequilibrium nondissipation principle, and is suitable for activated low-pressure diamond growth from the vapor phase.

8.3 Influences of *T* and *p* on Projective Phase Diagrams for C–H–O Systems

As mentioned above, a complete *T–p–X* phase diagram of a C–H–O ternary system must be in four-dimensional space. In the last section, the discussion focused only on a projective planar composition phase diagram for the 0.01–100 kPa pressure range and 700–1900 K temperature range. Obviously, projective planar composition phase diagrams for different pressure ranges and different temperature ranges must be different. Figure 8.4 shows projective phase diagrams for a fixed 0.01–100 kPa pressure range but for three different 700–1900 K, 900–1900 K and 1100–1900 K temperature ranges.[8, 14, 16, 17, 23, 28, 34] Bachmann's empirical phase lines are still included for comparison. From Fig. 8.4 it is clearly seen that the possible diamond growth phase region obviously shrinks with the increase of the lower temperature. In particular, close to the C–O sideline of the triangle the shrinkage is more remarkable. The lower carbon concentration phase line of the possible diamond growth region (i.e. the lower carbon concentration limitation for diamond growth) rises quickly close to the tie line of H–CO with an increase of the low-temperature limitation of the substrate temperature range. The higher carbon concentration line of the possible diamond growth region (i.e. the higher carbon concentration limitation for diamond growth) lowers with increase of the low-temperature limitation of the temperature range. However, the right end (on the C–O sideline of the triangle) of lines remain nearly unchanged at the fixed point of the CO position. In this way, the possible diamond growth phase region of the projective phase diagram for the 1300–1900 K temperature range will be nearly a very narrow tapered shaped triangle with an apex on the C–O sideline of the ternary phase diagram. The lower carbon concentration border is very close to the H–CO tie line. The narrow tapered shaped triangle for the 1300–1900 K temperature range is not drawn out in Fig. 8.4, but it can be found from the change tendency. In comparison with Bachmann's empirical phase region, such a theoretically possible diamond growth region is much narrower in shape than the empirical phase region. Moreover, the apex of the narrow triangle is just in the opposite direction. That clearly reflects the most important disadvantage of Bachmann's empirical phase diagram, which did not represent the influences of temperature and pressure. Such a disadvantage of Bachmann's empirical phase diagram has been uncovered more clearly by critical experiments, which will be discussed in the next section.

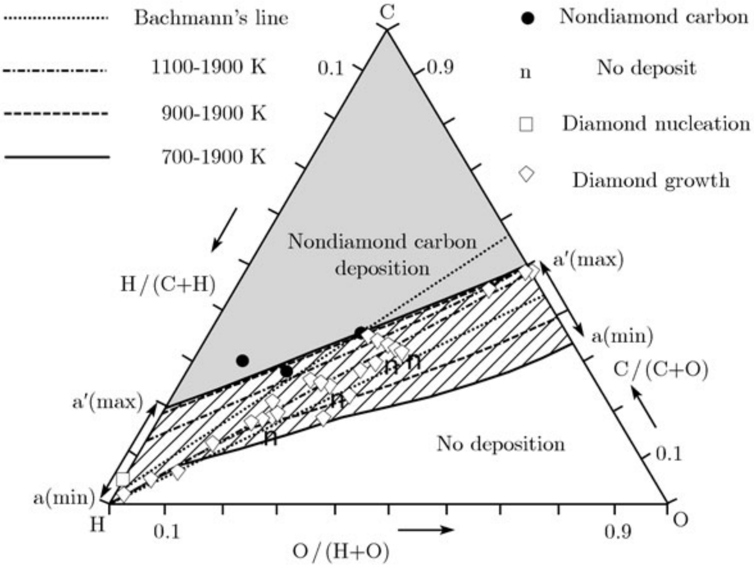


Fig. 8.4 Theoretically calculated projective nonequilibrium phase diagrams for activated low-pressure diamond growth from ternary C–H–O systems in different temperature ranges under 0.01–100 kPa

8.4 Marinelli's Critical Experimental Phase Diagram

In 1994, Marinelli et al. made a detailed comparison between Bachmann's empirical phase region and their critical experiments, as shown in Fig. 8.5.[40] Marinelli's results confirmed the existence of a single compositional "diamond domain" in a C–H–O triangular diagram. Bachmann correctly divided the triangular phase diagram into three phase regions, i.e. (i) no growth region, (ii) diamond growth region, also called the "diamond domain" and (iii) nondiamond carbon growth region. However, Marinelli et al. also concluded that the shape and the border of their experimental diamond growth region were quantitatively quite different from those indicated by Bachmann et al. According to Marinelli's study by the plasma optical emission spectra, the phase line between the no growth region and the diamond growth region was independent of the identity of the initial chemical species. That showed the presence of the same mechanisms of diamond growth

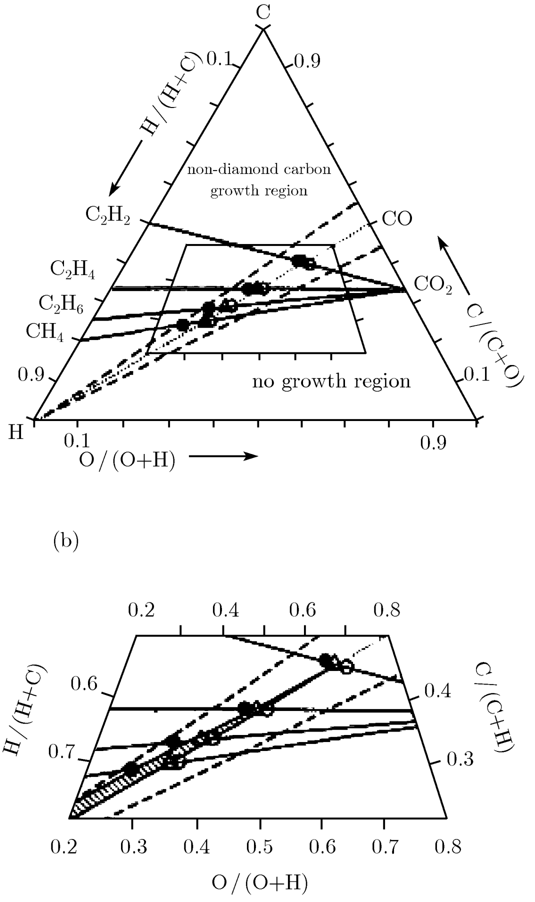


Fig. 8.5 Marinelli's original empirical phase diagram at 1173 K and 6.6 kPa. (○) no growth, (Δ) diamond, and (•) graphite deposition. Reprinted from M. Marinelli, E. Milani, M. Montuori, et al. (Figure 4, p. 5704)[40] by permission of American Institute of Physics

from the vapor phase for all the gas mixtures investigated.

The growth gas mixtures investigated were CH_4–CO_2, C_2H_6–CO_2, C_2H_4–CO_2 and C_2H_2–CO_2. The concentration of hydrocarbons increased gradually, so that the total composition of the system was consequent to moving along the corresponding tie lines of CH_4–CO_2, C_2H_6–CO_2, C_2H_4–CO_2 and C_2H_2–CO_2, respectively. Marinelli's experiments were run at a substrate temperature of 900 °C (1173K) and a pressure of 50 Torr (6.6 kPa). They found that (i) the composition phase line between the no growth region and the diamond growth region was extremely sharp; (ii) just over this phase line into the diamond growth region good quality diamond films were obtained, and SEM photos showed the perfection of crystals; (iii) moving deeply into the diamond growth region the diamond quality worsened and nonfaceted materials started appearing between the diamond crystallites; (iv) the phase line between the diamond growth region and the nondiamond carbon deposition region is rather fuzzy. Around this phase line the amount of nondiamond material increases continuously and the diamond facets are gradually lost.

Quantitatively, the diamond growth region deduced from Marinelli's experimental data considerably differs from Bachmann's one in that: (i) there exists a very sharp phase line between the no growth region (gas phase region) and the diamond growth region, and the line coincides (within experimental uncertainty) with the H–CO tie line itself; (ii) the width of the diamond growth region shrinks while moving away from the H apex. In the case of a C_2H_2–CO_2 mixture, Marinelli found only a single composition (C_2H_2 67.3 %, CO_2 32.7 %) which results in diamond deposition. Even deviations in gas composition of the order of 1 % will make the working point fall either in the no growth region or in the nondiamond carbon deposition region. The results obtained for all of the different gas mixtures are summarized in Table 8.1.

Fig. 8.1 Relationship between Marinelli's deposition results and hydrocarbon concentrations in CO_2 [40]

	CH_4	C_2H_6	C_2H_4	C_2H_2
no deposit	47.7	32.5	32.1	32.0
diamond	49.1	34.2	35.1	32.7
nondiamond carbon	56.3	39.4	37.4	33.4

8.5 Cross-Section Nonequilibrium Phase Diagrams for C–H–O Systems

Figure 8.6 is a theoretically calculated isothermal and isobaric cross-section nonequilibrium phase diagram at exactly the same temperature (900 °C, i.e. 1173 K) and pressure (6.6 kPa) as reported by Marinelli et al. in their critical experiments.[40] When comparing the projective nonequilibrium phase diagram in a narrow temperature range of 1100–1500 K in Fig. 8.4, it can be found that both of them are very similar. This is because of the narrow temperature range already used. In fact, if the projective ranges of temperature and pressure are shrunk into a certain temperature point and a certain pressure point, then the projective phase diagram will be exactly the same as the isothermal and isobaric cross-sectional phase diagram. In Fig. 8.6 Marinelli's critical experimental data had already been included, so the quantitative agreement between theory and experiments is shown very clearly. If a direct comparison between Fig. 8.6 and Fig. 8.5 is made, then theoretical phase diagram is more reasonable because the theoretical diamond growth region shrinks just at the C–O side line. On the other hand, if the apex of the diamond growth region in Marinelli's experimental phase diagram were a little extended to the right C–O sideline, the experimental phase diagram, Fig. 8.5, would be more reasonable, and exactly the same as our theoretical phase diagram, Fig. 8.6. These results confirmed either that Marinelli's experiments are really critical experiments with very good accuracy or that our theoretical nonequilibrium phase diagram agrees excellently with critical experiments.

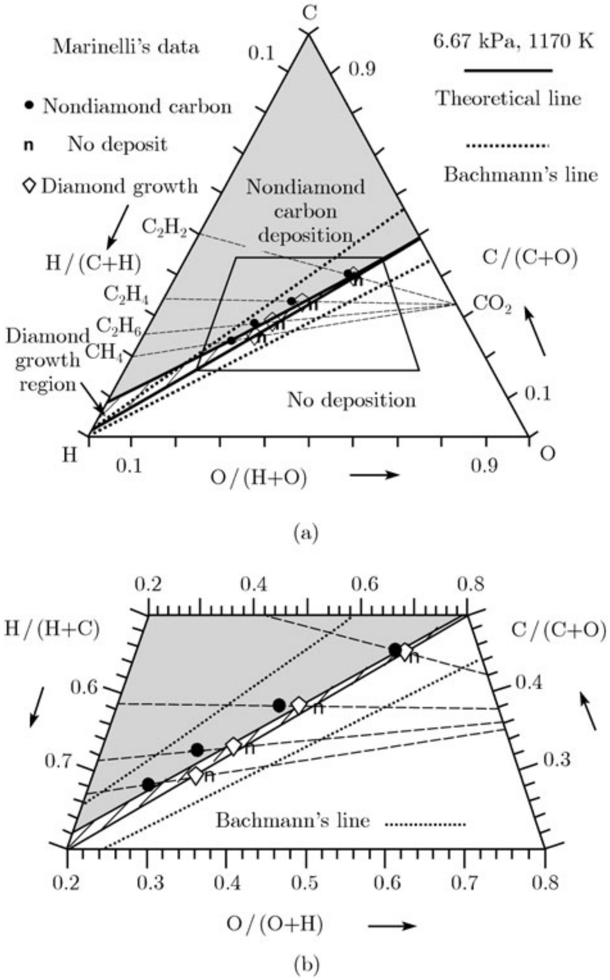


Fig. 8.6 Excellent agreement between theoretical cross-sectional nonequilibrium phase diagram and Marinelli's critical experiments [40]. (a) Cross-sectional nonequilibrium phase diagram calculated at 1173 K and 6.6 kPa. (b) Amplified phase diagram of the ladder-shaped part in (a)

Due to the complexity of T–p–X phase diagrams of ternary systems, phase diagrams of a four-dimensional character can also be studied by way of two-dimensional isothermal isobaric cross-section phase diagrams. It should be noted that in these isothermal isobaric cross sectional phase diagrams all variables of temperature, pressure and composition are completely fixed. There is no possibility of finding an overlap between different phase regions in each of these isothermal isobaric cross-sectional phase diagrams. The overlap between different phase regions usually occurs in projective phase diagrams. Therefore, the names of phase regions here are: (i) no deposition region (i.e. gas phase region), (ii) diamond growth region and (iii) nondiamond carbon deposition region. Among them phase region (ii) is called the diamond growth region in cross-sectional phase diagrams, while the similar phase region (ii) is called the possible diamond growth region in projective phase diagrams. It means that if the experimental conditions (i.e. composition, temperature and pressure of the system) are inside the diamond growth region of the isothermal isobaric cross-section phase diagrams, the thermodynamic possibility is only for diamond growth, but not for no growth or nondiamond carbon deposition.

From the discussion above in this chapter it is clearly known that the nature of Bachmann's phase diagram is an empirical projective nonequilibrium phase diagram. The corresponding ranges of projections were (>400 °C, 0.1–760 Torr) or close to the ranges selected in our theoretical calculations, i.e. 700 K–1900 K and 0.01 kPa 100 kPa. Therefore, Bachmann's diamond growth region should include all reliable experimental data of diamond growth. Because there should be only a "possible diamond growth region", the interleaving of "no deposition" and "nondiamond carbon deposition" is possible and reasonable. At the same time, it is clearly known that the nature of Marinelli's phase diagram is an empirical cross-section nonequilibrium phase diagram. Of course, it is also an empirical isothermal isobaric cross-section nonequilibrium phase diagram at a fixed temperature (900 °C or 1173 K) and a fixed pressure (50 Torr or 6.6 kPa). If its diamond growth region were extended to the C–O side line, Marinelli's phase diagram would be a perfect cross-section nonequilibrium phase diagram. Therefore, Bachmann's empirical phase diagram and Marinelli's empirical phase diagram are not to be compared with each other. Bachmann's empirical phase diagram and Marinelli's empirical phase diagram should be compared with the corresponding theoretical projective and cross-section nonequilibrium phase diagrams, respectively, to get confirmations and reach a new level of recognition.

8.6 Nonequilibrium Phase Diagrams for C–H–X Systems

Similar to the calculation method used for the C–H–O system, nonequilibrium phase diagrams of C–H–F and C–H–Cl ternary systems had also been calculated for activated low-pressure diamond growth from the vapor phase, as shown in Fig. 8.7, and Fig. 8.8, respectively.[41~48] The experimental data reported in the literature are also illustrated. The agreement between theoretically calculated results and experimental data is also very nice.

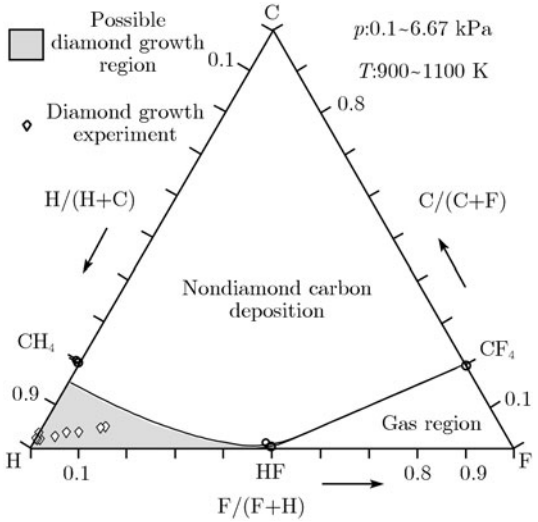


Fig. 8.7 Comparison between calculated projective nonequilibrium phase diagrams and experiments for diamond growth from C–H–F systems

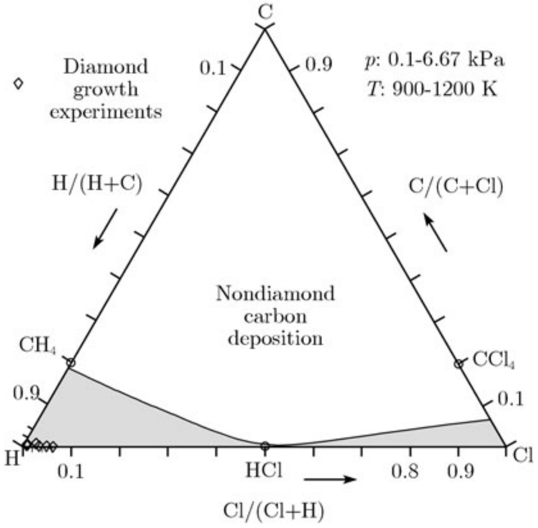


Fig. 8.8 Comparison between calculated projective nonequilibrium phase diagrams and experiments for diamond growth from C–H–Cl systems

8.7 Nonequilibrium Phase Diagrams for Low-Pressure cBN Syntheses

Due to the similarities between allotropic pairs of diamond–graphite and cubic boron nitride–hexagonal boron nitride (cBN–hBN), the new kind of nonequilibrium phase diagram can also be used for activated CVD cBN processes. Figure 8.9 and Fig. 8.10 are projective nonequilibrium phase diagrams calculated on the basis of our coupling model for activated (by hot filament, plasma torch, and/or r.f. plasma, $\cdots$) cBN growth from BN–H–Cl and BN–H–F systems, respectively, in the range of total pressure $p = 10–10^5$ Pa and substrate temperature $T = 1200–1500$ K.[49~51] There are three typical phase regions: (i) gas region or no deposition region, (ii) possible cBN growth region or cBN stable region (filled by oblique lines), and (iii) non-cBN deposition region (filled by net-like pattern). One must keep the total composition of the system inside the cBN stable region for growing cBN crystals or films. However, the possible cBN growth region is much narrower in comparison with the diamond growth region in the corresponding phase diagrams for the activated CVD diamond process. The possible cBN growth region seems to be a line in both Fig. 8.9 and Fig. 8.10, but it is really a region, as shown in the small amplified circles.

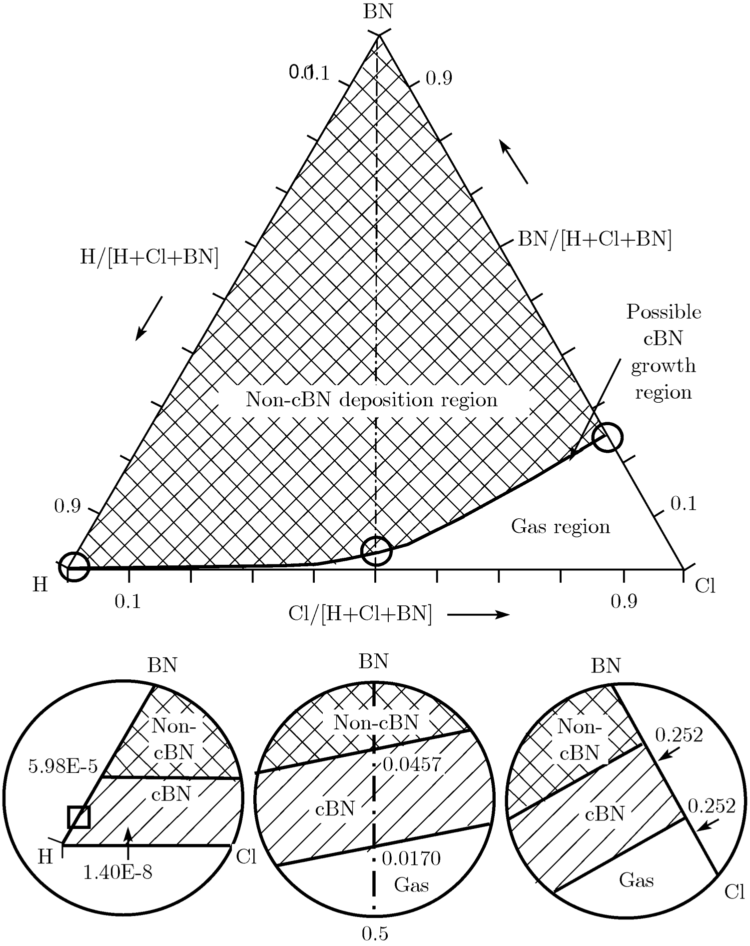


Fig. 8.9 Nonequilibrium phase diagram of BN–H–Cl system for activated CVD of cBN in the range $p = 10–10^5$ Pa and $T = 1200–1500$ K. The highest and lowest BN/(H+Cl+BN) values for cBN growth are shown amplified in the three small circles. The open square mark on the BN–H sideline is Song's experimental point[53] inside the theoretical range of BN/(H+BN) = 1.40×10^{-8}–5.98×10^{-5}

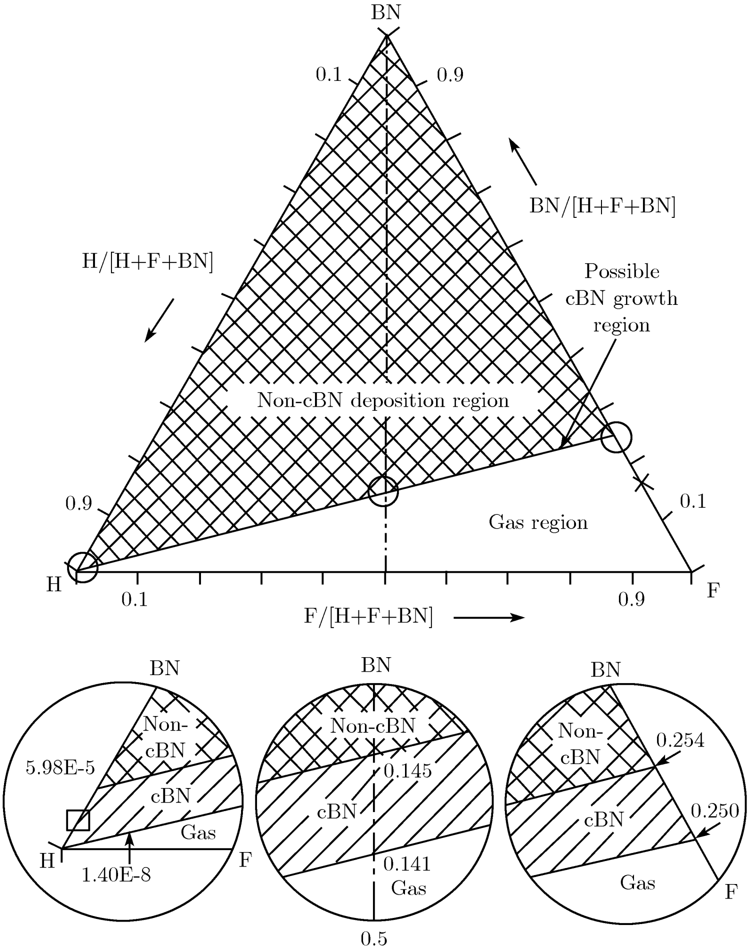


Fig. 8.10 Nonequilibrium phase diagram of BN–H–F system for activated CVD of cBN in the range p = 10–10^5 Pa and T = 1200–1500 K. The highest and lowest BN/(H+F+BN) values for cBN growth are shown amplified in the three small circles. The open square mark on the BN–H sideline is Song's experimental point [53] inside the theoretical range of BN/(H+BN) = 1.40×10^{-8}–5.98×10^{-5} for cBN growth; but the mark × on the BN–F sideline is Matsumoto's unsuccessful experimental point [54]

That is perhaps the reason why cBN is much more difficult to prepare under low pressure, although activated CVD processes for cBN are very similar to those for diamond. For instance, the hot filament process has been used very easily and stably for the activated CVD diamond process, but a similar hot filament process must be enhanced by a simultaneous r.f. plasma or bias for cBN growth, as reported by F. Zhang, Z. Song, Y. Guo and G. Chen. [52, 53] Figure 8.9 and Fig. 8.10 also show that the cBN growth region from a BN–H system is very close to the apex of H, so cBN can be obtained from B_2H_6–NH_3 gas sources diluted by a large amount of hydrogen (the ratio of B_2H_6 : H_2 is about 0.05: 2400, i.e. 2.1×10^{-5} marked by an open square, and our theoretical value is 1.40×10^{-8}–5.98×10^{-5} in Fig. 8.9 and Fig. 8.10) with activation by the hot filament together with the r.f. plasma, as reported by Song et al.[53] Any attempt with a much higher concentration of B_2H_6–NH_3 in H_2 will get hBN or non-cBN instead of cBN. This is because the gas composition comes into the non-cBN deposition region. That explains Matsumoto's experiment.[54] They got non-cBN films from a $B_2H_6 : NH_3 : H_2$ (2.5 : 10 : 1000) mixture (although diluted by Ar) because the concentration of about 0.25 % B_2H_6 in H_2 was too high in comparison with our theoretical value of 1.40×10^{-8}–5.98×10^{-5}, and came into the non-cBN deposition region in Fig. 8.9 and Fig. 8.10. For the BN–F system, there is no possibility of getting cBN from the BF_3–NF_3 system because the BN: F ratio is fixed at 1: 6 (marked by ×) and comes into the gas region on the BN–F sideline of Fig. 8.10. It had been reported by Matsumoto et al.[54] that "The formation of cubic BN was not confirmed" from the BF_3–NF_3 system even with strong activation of a plasma torch together with another r.f. plasma. If any kind of BN films from a "BF_3–NF_3" gas mixture were obtained, that was due to the reduction of the substrate (Si or Mo), or other reductants inside the reaction chamber. The BN–H–O phase diagram for the activated low-pressure cBN growth wasn't calculated here, although C–H–O phase diagrams are very important for activated low-pressure diamond growth. That is due to the formation of B_2O_3, if any kind of oxygen (including chamber leakage) comes into the reaction chamber. B_2O_3 could form, liquefy and cover the substrate to prevent the growth of cBN.

The existence of the cBN growth region in Fig. 8.9 and Fig. 8.10 shows that cBN has become a stable phase under activated low-pressure conditions, while hBN has been activated by superequilibrium atomic hydrogen (denoted by H*) [together with superequilibrium atomic chlorine (Cl*), or superequilibrium atomic fluorine (F*)] into a metastable phase. For instance, on the basis of thermodynamic coupling reactions, as shown in reaction Eq. (8.1) (denoted by subscript 1), reaction Eq. (8.2) (denoted by subscript 2), and reaction Eq. (8.3) (denoted by subscript 3), thermodynamic data for H* activated hBN, denoted by hBN*(H), could be calculated. The calculated data for hBN*(H) with the rate ratio of reaction Eq. (8.2) to reaction Eq. (8.1) (denoted by χ) are listed in Table 8.2.

Fig. 8.2 Thermodynamic data for hBN*(H) with χ = 0.28

$H_{\mathrm{hBN*}}(\mathrm{H}) = -189.464\ \mathrm{kJ\ mol^{-1}}$			
$S_{\mathrm{hBN*}}(\mathrm{H}) = 29.143\ \mathrm{J\ mol^{-1}\ K^{-1}}$			
$C_p\ (\mathrm{J\ mol^{-1}\ K^{-1}}) = A + B10^{-3}T + C10^5T^{-2} + D10^{-6}T^2$			
	1550 K	1550–1750 K	1750–3000 K
A	19.5121	19.5648	50.6139
B	40.2709	39.4659	−0.4578
C	−8.0777	−2.3693	5.7920
D	−13.0917	−12.6859	0.0301

$T_{\mathrm{activation}}$: 3000 K, Ref. T: 298 K, p range: 10–10^5 Pa

Definitions of B, C and D here are a little different from those in Chap. 3

$$\mathrm{hBN} = \mathrm{cBN} \qquad (8.1)$$
$$(1400\mathrm{K}, 100\mathrm{Pa}),$$
$$\Delta G_1 > 0.$$

$$\mathrm{H}{*} = 0.5\mathrm{H}_2 \tag{8.2}$$
$$(1400\mathrm{K}, 100\mathrm{Pa}, T_{\text{activation}} = 3000\mathrm{K}),$$
$$\Delta G_2 < 0.$$

Eq. (8.3) = Eq. (8.1) + χ Eq. (8.2),

$$\mathrm{hBN} + \chi\mathrm{H}{*} = 0.5\chi\mathrm{H}_2 + \mathrm{cBN} \tag{8.3}$$
$$(1400\mathrm{K}, 100\mathrm{Pa})$$

or

$$\mathrm{hBN} * (\mathrm{H}) = [\mathrm{hBN} + \chi(\mathrm{H}{*} - 0.5\mathrm{H}_2)] = \mathrm{cBN} \tag{8.4}$$
$$(1400\mathrm{K}, 100\mathrm{Pa}),$$
$$\Delta G_3 = \Delta G_4 < 0, \mathrm{if}\chi \geqslant 0.28.$$

Due to this similarity, thermodynamic data for Cl* and F* activated hBN will not be discussed and listed here. For simplicity, the fixed ratio of B/N = 1 is used during calculation. In practice, a certain excess of N-containing precursors (such as NH_3, NF_3, $\cdots$) is usually used.

It should be noted that all other standard thermodynamic data could be found in ordinary data books, mainly from JANAF, [55] but the entropy (S) and constant pressure heat capacity (C_p) of cBN are taken from another reference (V.L. Solozhenko). [56, 57] Usually, there is no doubt about the standard Gibbs free energy of hBN (at 298 K and 10^5 Pa, $G^0_{\mathrm{hBN}} = -254.91 \pm 1.5$ kJ mol^{-1}), but there were different values for the standard Gibbs free energy of cBN (G^0_{cBN}). However, up to now cBN spontaneous crystallization has not been confirmed below 2.0 GPa. Suppose that the equilibrium pressure between hBN and cBN at 298 K is 1.5 GPa, i.e. the Gibbs free energy change of the phase transformation from hBN to cBN $\Delta G(p) = 0$ (at 298 K and 1.5 GPa), based on the relation between pressure and the Gibbs free energy, $\left(\frac{\partial \Delta G}{\partial p}\right)_T$, one can get

$$\begin{aligned}\Delta G(p) - \Delta G(p^0) &= \int_{p^0}^{p} \Delta V \mathrm{d}p = \Delta V(p - p^0) \\ &= \left\{ \left(\frac{24.83}{3.48} - \frac{24.83}{2.29} \right) \times 10^{-6} \mathrm{m}^3 \mathrm{mol}^{-1} \right\} \\ &\times \left(1.5 \times 10^6 - 101.325\right) \mathrm{kPa} \\ &= -5.55 \mathrm{kJmol}^{-1}.\end{aligned} \tag{8.5}$$

In this way, the standard Gibbs free energy of cBN (G^0_{cBN}) at 298 K equals -249.36 kJ mol^{-1} here. In Eq. (8.5), 24.83, 3.48 g cm^{-3} and 2.29 g cm^{-3} are the molecular weight of BN, the densities of cBN at 298.15 K, and the densities of hBN at 298.15 K, respectively. For χ it is known that in the activated cBN process a much stronger activation condition is needed than that in the activated diamond process, and that χ may be about 0.28 or bigger in the diamond process, so a lower limitation of χ = 0.28 used for calculation here is reasonable. The discussion on the activated low-pressure cubic boron nitride shows that the new field of nonequilibrium nondissipative thermodynamics is not only useful for the low-pressure diamond process, but also for other thermodynamics problems in new materials and new technology. Especially, a lot of projects in life sciences need new theory of modern thermodynamics.

8.8 Evaluations and Brief Summary on Nonequilibrium Phase Diagrams

In regard to nonequilibrium phase diagrams of the activated low-pressure diamond growth, in 2001, B. Heimann, V. Raiko & V. Buck in *Refractory Metals & Hard Materials* cited a series of our 5 papers published during 1995–1997 [24, 25, 32, 35, 36] and evaluated: [58]

> To answer the question, why diamond can grow under low-pressure conditions, the thermodynamic coupling model proposes a nonequilibrium thermodynamic activation. The driving force for the transformation from graphite state of carbon, stable in thermodynamic equilibrium, to diamond under low pressure comes from a coupled association reaction of superequilibrium of the atomic hydrogen. The concentration of H is in equilibrium at the hot-filament temperature, but in superequilibrium at the much lower substrate temperature. The projection of phase diagrams obtained from the thermodynamic coupling model may be crucial for understanding Bachmann's diagram.

In China an evaluation in 2000 also said: [59]

> There were some creations in China on the researches of diamond films, such as 'thermodynamic coupling model of the activated low-pressure diamond growth' and the entirely new concept of 'stationary nonequilibrium phase diagrams' were proposed firstly in the international forums. All of these progresses were focused attention upon.

In 2001, a book on low-pressure diamonds written by D.-H. Dai and K.-S. Zhou clearly pointed out: [60]

> Comparing with Bachmann's empirical C–H–O ternary phase diagram, $\cdots$ the calculated C–H–O ternary stationary nonequilibrium phase diagrams on the basis of nonequilibrium thermodynamics are much better in the coverage of experimental data and academic significances. Authors believe that the nonequilibrium phase diagrams are more perfect and their instructions on the technology are of more strong persuasions.

and so on. In 2004, there were also some similar evaluations in publication of D.-H. Dai, K.-S. Zhou, Z.-H. Yuan et al. [61] The author will not repeat them here.

To sum up, in modern thermodynamics the continuous stable diamond growth under the activated low-pressure conditions is a normal thermodynamic phenomenon. In other words, the low-pressure diamond growth can be completely realized under supply of external energy in several forms, such as providing and maintaining a superequilibrium concentration of atomic hydrogen via hot filament or plasma activation. In different processes, the concentrations and the association of superequilibrium atomic hydrogen may be quite different. If the activation of external energy is stronger, the stable diamond growth region will be broader, and the low-pressure diamond growth will be easier.

Finally, it should still be noted that due to the instability of diamond bulks or crystals at very high temperature (maybe 1700 K, or in different references with different conditions 1900 K), thermodynamic data of diamond are usually valid below 1200 K. Thermodynamic data of diamond over 1200 K are only extended results. The theoretical calculated lines should be taken as references in predictions. However, based on the nonequilibrium thermodynamic coupling model (or coupling model, for short) for the activated low-pressure diamond growth from the vapor phase, a new concept of nonequilibrium phase diagrams has been proposed. Nonequilibrium phase diagrams of carbon-hydrogen systems during the existence of superequilibrium concentration of atomic hydrogen have been rather strictly calculated. Nonequilibrium phase diagrams may be used as a theoretical guide not only for improvement of low-pressure diamond process, but also for exploiting new materials and their processes on the basis of modern thermodynamics.

To sum up the last two chapters, the significance of nonequilibrium phase diagrams is due to at least one more dimension of nonequilibrium parameters having been introduced into calculation of phase diagrams (CALPHAD). This is a new CALPHAD of modern thermodynamics.

References

1. Bachmann P K, Leers D, Lydtin H. Diamond and Related Materials [J], 1, 1-12 (1991)
2. Bachmann P K, van Enckevort W. Diamond and Related Materials [J], 1, 1021-1034 (1992)
3. Wang J-T. Modern thermodynamics – A whole view of thermodynamics [M]. Shanghai: Fudan Univ Press, 2005.
4. Wang J-T. Nonequilibrium Nondissipative Thermodynamics — With Application to Low-Pressure Diamond Synthesis [M]. Heidelberg: Springer, 2002.
5. Wang J-T. Phase Diagrams of Stationary Nonequilibrium States — Thermodynamics for Low Pressure Diamond Growth from the Vapor Phase [M]. Beijing: Science Press, 2000, 212 pages.
6. Wang J-T, Zhang D W, Liu Z-J. Thermodynamic Coupling Model for Low Pressure Diamond Growth from the Vapor Phase [M]. Beijing: Science Press, 1998, 163 pages; reprinted in 2000, 175 pages.
7. Wang J-T, Zhang D W, Ding S-J, Wang P-F. CALPHAD (J of Calculated Phase Diagrams) [J], 2000, 24: 427-434.
8. Liu Z-J, Zhang D W, Wang P-F, Ding S-J, Zhang J-Y, Wang J-T. Thin Solid Films [J], 2000, 368(2): 253-256.
9. Wang J-T. Physics (Beijing) [J], 2000, 29(9): 524-530.
10. Wang J-T. Modern thermodynamics for phase diagrams of stationary states (plenary lecture). In: Phase Diagram Calculation (CALPHAD XXIX) [C], MIT, Cambridge, Massachusetts: 2000, 8.
11. Wang J-T. Ding S-J, Zhang D W, Zhang J-Y, Wang P-F. Calculation of phase diagrams of stationary states. In: Chemical Thermodynamics (ICCT-2000) [C], Halifax, Canada: 2000, 49.
12. Wang J-T. In: Materials Research (CMRS-2000), Abstracts of 2000 CMRS Meeting [C], Beijing: Oct 2000, v 1: 438.
13. Wang J-T. In: Proc of 10th National Symp on Phase Diagram [C], Chinese Physical Society, Wuhu, Anhui: 25-28 Sep 2000, 135-138.

14. Liu Z-J, Zhang D W, Wan Y-Z, Zhang J-Y, Wang J-T. Chinese Chemistry Bulletin [J], 1999, 1999(2): 59-62.
15. Wang J-T, Zhang D W. From activated CVD diamond to the life science (invited talk). In: Chemical Vapor Deposition, Abstracts of 1999 Asian Conference [C], 10-13 May 1999, Shanghai, paper no 1-6
16. Wang J-T, Wan Y-Z, Liu Z-J, Wang H, Zhang D W, Huang Z-Q. Materials Letters [C], 1998, 33: 311-314.
17. Wan Y-Z, Zhang D W, Liu Z-J, Wang J-T. Applied Physics A [J], 1998, 67: 225-231.
18. Wan Y-Z, Zhang D W, Wang J-T. J Applied Science [J], 1998, 16(2): 185-190.
19. Wan Y-Z, Zhang D W, Liu Z-J, Wang J-T. J Synthetic Crystals [J], 1998, 27(2): 109-113.
20. Liu Z-J, Zhang D W, Wan Y-Z, Wang J-T. Functional Materials [J], 1998, 29(5): 506-508.
21. Wang J-T. Physics [J], 1998, 27(2): 77-83.
22. Wang J-T, Zhang D W, Liu Z-J, J-Y Zhang. Structure stability in plasma chemistry (invited). In: Plasma Deposition and Treatment of Polymers, Symp of ’98 Fall Meeting of Materials Research Society [C], at Boston, 30 Nov - 2 Dec 1998, ed by Lee W W, d’Agostino R, Wertheimer M R, Warrendale: Materials Research Society, 1998, SPV544: 101-108.
23. Wang J-T, Zhang D W, Liu Z-J, Wan Y-Z, Zhang J-Y, 27th Internat Conf on Phase Diagrams and its Application [C], Beijing: May, 1998.
24. Wang J-T, Huang Z-Q, Wan Y-Z, Zhang D W, H-Y Jia, J of Materials Research [J], 1997, 12(6): 1530-1535.
25. Wang J-T, Wan Y-Z, Zhang D W, Liu Z-J, Huang Z-Q. J Materials Research [J], 1997, 12(12): 3250-3253.
26. Wang J-T, Huang Z-Q, Wan Y-Z, Zhang D W, Yang Q-H. Chinese Science Bulletin [J], Eng ed. 1997, 42(11): 967-968.
27. Wang J-T, Huang Z-Q, Wan Y-Z, Zhang D W, Zheng P-J. Progress in Natural Science [J], Eng Edn. 1997, 7(3): 265-271.
28. Wan Y-Z, Zhang D W, Liu Z-J, Wang J-T. High Technol Lett [J], 1997, 3(2): 80-83.
29. Wang J-T, Wan Y-Z, Liu Z-J, Zhang D W, Huang Z-Q. Comparison of phase diagrams with experiments of activated diamond growth from the vapor phase. Presented at 5th International Symposium on Diamond Materials [C], Paris, France: 31 Aug - 5 Sep 1997.
30. Wang J-T, Wan Y-Z, Liu Z-J, Zhang D W. In: Phase Diagrams, 9th National Symp at Beijing [C], 21-23 Oct 1997, ed by Chinese University of Geosciences, and Commission of Phase Diagrams of Chinese Physical Society, 17-19.
31. Wan Y-Z, Wang J-T, Liu Z-J, Zhang D W. In: Phase Diagrams, 9th National Symp at Beijing [C], 21-23 Oct 1997, ed by Chinese University of Geosciences, and Commission of Phase Diagrams of Chinese Physical Society, 4-6.
32. Wang J-T, Huang Z-Q, Yang Q-H, Zhang D W, Wan Y-Z, In: Proc of the 13th Inter Conf on Chemical Vapor Deposition [C], eds Besmann T M, Allendorf M D, Robinson McD, et al. Pennington, NJ: The Electrochem Soc, 1996, PV96-5: 727-732.
33. Wang J-T, Zhang D W, Wan Y-Z, Huang Z-Q. Presented as paper no. 8.092b at Diamond ’96 Conf [C], Tours, France: 8-13 Sep 1996.
34. Wan Y-Z, Zhang D W, Wang J-T, Huang Z-Q. Presented at ’96 CMRS Meeting [C], Section C Beijing, ’96 Proc of Chinese MRS, Functional Materials (I-3) Beijing: Chemical Industry Press, 1997, 552-555.

35. Wang J-T, Zheng P-J, Yang Q-H, Wang H. In: Proc of the 4th Internat Symp on Diamond Materials [C], eds Ravi K V, Dismukes J P. Pennington, NJ: The Electrochem Soc, 1995, PV95-4: 13-19.
36. Wang J-T, Zheng P-J. Chinese Science Bulletin [J], Eng edn. 1995, 40(13): 1141-1143.
37. Mao M-Y, Jin X-F, Wang T-P, et al. Appl Phys Lett [J], 1995, 66: 16.
38. Bachmann P K, Hagemann H J, Lade H, et al. In: Diamond, SiC and Nitride Wide Bandgap Semiconductor [C], Spring Meeting of Materials Research Society at San Francisco, California 1994, ed by Carter Jr C H, Gildenblat G, et al. 1994, 267-277.
39. Bachmann P K, Hagemann H J, Lade H, et al. Diamond and Related Materials [J], 1995, 4: 820-826.
40. Marinelli M, Milani E, Montuori M, et al. J Appl Phys [J], 1994, 76(10): 5702-5705.
41. Liu Z-J, Zhang D W, Y-Z Wan, Wang J-T. Applied Physics A [J], 1999, 68(3): 359-362.
42. Liu Z-J, Zhang D W, Zhang J-Y, Wan Y-Z, Wang J-T. Thin Solid Films [J], 1999, 342: 42-46.
43. Liu Z-J, Zhang D W, Zhang J-Y, Wan Y-Z, Wang J-T. Progress in Natural Science [J], Chinese edn. 1999, 9(4): 378-380.
44. Wan Y-Z, Zhang D W, Liu Z-J, Zhang J-Y, Wang J-T. Materials Chemistry and Physics [J], 1998, 56: 275-279.
45. Wan Y-Z, Zhang J-Y,, Liu Z-J, Zhang D W, Wang J-T. Chinese J of Materials Research [J], 1998, 12(6): 604-609.
46. Wan Y-Z, Liu Z-J, Zhang D W, Wang J-T. Microfabrication Technology [J], 1997, 4: 43-48.
47. Liu Z-J, Zhang D W, Wan Y-Z, Wang J-T. J Inorganic Materials [J], 1998, 13(4): 634-636.
48. Liu Z-J, Zhang D W, Zhang J-Y, Wan Y-Z, Wang J-T. J Synthetic Crystals [J], 1998, 27(3): 201-205.
49. Wang J-T, Zhang J-Y, Ding S-J, Zhang D W, Wang P-F. Phase diagrams for activated CVD diamond and cubic BN. In: Chemical Thermodynamics (ICCT-2000) [C], Abstracts of the 16th IUPAC International Conf, Halifax, Canada: 6-11 Aug 2000, 43.
50. Wang J-T, Ding S-J, Zhang D W, Zhang J-Y, Wang P-F. In: Internat Conf on Eng & Tech Sci 2000 [C], ed by Song J, Yin R-Y. Beijing: New World Press, 2000, 857-863.
51. Wang J-T, Shen J-Y, Zhang D W. International J Refractory Metals and Hard Materials [C], 2001, 19(4-6): 461-466.
52. Zhang F, Guo Y, Song Z, Chen G. Appl Phys Lett [J], 1994, 65(8): 971.
53. Song Z, Zhang F, Guo Y, Chen G. Appl Phys Lett [J], 1994, 65(21): 2669.
54. Matsumoto S, Nishida N, Akashi K, Sugai K. J of Mater Sci [J], 1996, 31: 713.
55. Stull D R, et al. JANAF Thermochemical Tables [DB]. 2nd edn. Washington: 1971; Chase Jr M W, Davies C A, Downey Jr J R, et al. JANAF Thermochemical Tables [DB], 3rd ed. J Phys Chem Ref Data [J], 1985, 14 Suppl.
56. Solozhenko L. Diamond and Related Materials [J], 1994, 4: 1-4.
57. Solozhenko L. J Hard Materials [J], 1995, 6: 51-65.
58. Heimann B, Raiko V, Buck V. Refractory Metals & Hard Mater [J], 2001, 19(3): 169.
59. Engineering Academy of China, Academy of China, Status Quo of Materials Development of China and Countermeasure for Going to the New Century, Consultation Program. Consultation Group on Nonnonferrous Metallic Materials (ed). Consultation Report on Nonnonferrous Metallic Materials [R]. Shanxi: Shanxi Science and Technology Publisher, 2001, 229.
60. Dai D-H, Zhou K-S (eds). Preparation Technology and Applications of Diamond Film Deposition [M]. Beijing: Metallurgy Industry Publisher, 2001, Chap 2 and 191.
61. Dai D-H, Zhou K-S, Yuan Z-H et al (eds). Surface Science and Technology of Modern Materials [M]. Beijing: Metallurgy Industry Publisher, 2004, 372-373.

Chapter 9

Carat-Size Low-Pressure Diamonds and Other Thermodynamic Issues

Abstract In about 1970 the achievement of the activated chemical vapor deposition (CVD) diamond synthesis under low-pressure had been regarded as "alchemy" and thermodynamically impossible, although it is the case that only diamond films and tiny diamond crystals were obtained at that time. After about 30 years, however, beautiful 10-carat-sized, gem-quality single-crystal synthetic diamonds can already be obtained by the similar process under low-pressure. Such a thermodynamics enigma should still be explained by the thermodynamic coupling model which was proposed by us in 1990. This model arose as a result of the severe restriction of classical thermodynamics based on the Carnot theorem and Clausius' cyclical-process assumption. At the same time, Clausius' "compensation" (now called thermodynamic coupling) for complex systems, including biological processes and living systems, comes to our attention in the current 21st century. The existing living systems do not only agree with nature laws but also skillfully utilize natural laws, including the second law of thermodynamics. Finally, modern thermodynamics based on the extended Carnot theorem (or the nonequilibrium nondissipative principle) appears.

9.1 Carat-Size Gem-Quality Low-Pressure Diamond Growth

At *the* 10th *International Conference on New Diamond Science and Technology*, Tsukuba, Japan, on 12 May 2005, Russell Hemley and Ho-kwang Mao of the Carnegie Institution of Washington[1] announced the production of 10-carat, transparent colorless single-crystal diamonds at rapid growth rates (100 micrometers per hour) using a chemical vapor deposition (CVD) process. Also, at *the Applied Diamond Congress* in Argonne, Illinois, 18 May 2005, they made a presentation with the title "*Very Large Diamonds Produced Very Fast*".[1~5] The news then quickly spread through the media. Since about 1970, former Soviet scientists have grown diamond under low pressures by chemical vapor deposition (CVD). Diamond films and tiny diamond crystals with micrometers sizes were obtained, and the obtained low-pressure diamonds were usually with dark yellow, brown or even black color. These diamonds were impossible to be used for the preparation of jewellery. Now, the fabrication of 10-carat, half-inch, CVD diamonds in Carnegie is a major breakthrough not only for low-pressure diamond technology but also for the practical use of

---

[1] Ho-kwang Mao, A member of National Academy of Science, U.S.A. (since 1993). He is one of the highest pressure (over that in the center of the earth) record holders, and laureate of some international prizes

synthetic diamonds. It is well known that the biggest commercially available diamonds produced by the standard high-pressure/high-temperature (HPHT, > 4.5 GPa, $\approx 1200\,^{\circ}\mathrm{C}$) method are less than 3 carats (0.6 gram, about 7 mm diameter). In addition, the most HPHT synthetic diamonds are yellow and the most activated CVD diamonds were brown. This limited their applications, especially for optical applications. The team in Carnegie has made colorless single-crystal diamonds, transparent from the ultraviolet to infrared wavelengths, with their high growth rate CVD process. Of course, the real potential value of gem-quality diamonds is in relation to its extraordinary characteristics in optics and in scientific research. For instance, the rigidity and hardness is the highest of all known materials, so it can be used as anvil in high pressure researches. That was the original purpose of the Carnegie Institution. Diamond is important due to having the highest room temperature heat conductivity, the low heat expansion coefficient, an anti-radiation ability, chemical inertness to most acids, alkalis, and so on. Diamonds can also be used for fabrication of anti-radiation high-temperature semiconductor devices etc. Figure 9.1–Fig. 9.3 are photos of low-pressure carat-size diamonds in Carnegie.[2]

At the beginning of the 1990s the activated low-pressure diamond growth from the vapor phase was a focus of research projects across the world. At that time the growth rate of diamond film was about 1–10 μm per hour, while the growth rate of epitaxial layer on single-crystal diamond (for single-crystal diamond growth) was even slower (about 3 months for 1 mm diamond growth). Many researchers therefore stopped or gradually gave up their low-pressure diamond projects. However, Chih-shiue Yan and others at the Carnegie Institution of Washington persisted with their experiments. In 2002, the chemical vapor deposition (CVD) process had been successfully used by them for the diamond

Fig. 9.1 A brilliant cut single-crystal diamond produced from activated low-pressure chemical vapor deposition process [placed on mirror]. Reprinted from S.S. Ho, C.S. Yan, Z. Liu et al. [1] with permission from the Carnegie Institution of Washington

---

[2] http://www.carnegieinstitution.org/diamond13may2005/. Cited 25 Oct 2008

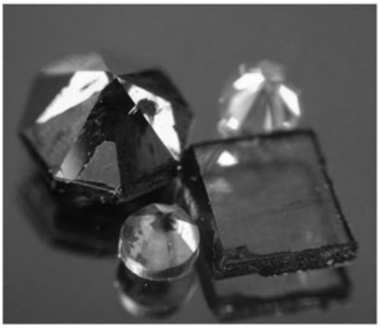


Fig. 9.2 The variety of single crystal diamonds produced by the Carnegie high-growth rate CVD process. Reprinted from website of the Carnegie Institution of Washington with permission from the Carnegie Institution of Washington

epitaxial growth. The new diamond epitaxial layer could be continuously grown on the natural diamond surface in the ambiance of hydrogen, nitrogen and methane mixture at high temperature. The growth rate was 300 μm per hour.[4] However, the color of obtained diamonds was not good at that time, having a yellow color. Now colorless single-crystal diamonds, transparent from the ultraviolet to infrared wavelengths have been made with the Carnegie rapid CVD process. If diamonds could be fabricated on a large scale in the future, then diamonds can be widely used as materials for incision, abrasion, heat conductor, medical components and electronics.

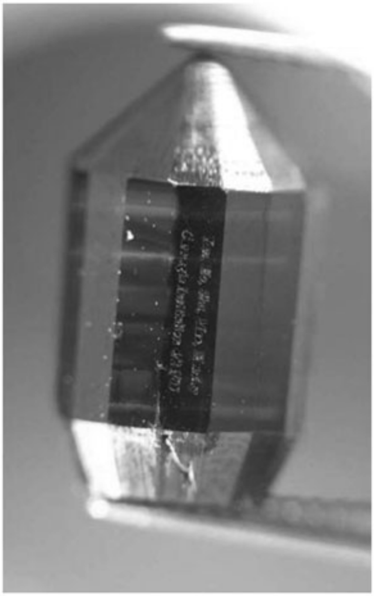


Fig. 9.3 Five carat single-crystal CVD diamond (without seed) 12 mm high and 6.7 mm in diameter cut from a 10 carat (8 x 8 x 12 mm^3) block. Reprinted from S.S. Ho, C.S. Yan, Z. Liu et al.[1] with permission from the Carnegie Institution of Washington

A few years ago, after the Carnegie patent application was published, hundreds of laboratories around the world came back to CVD diamond researches, but due to the failure of the growth rate, they withdrew from the competition. However, the Carnegie Institution spent millions of dollars and at last found the improvement for the growth rate and quality.

On one occasion they brought a low-pressure CVD grown diamond to a senior jewellery expert for identification. The expert said that the value should be about two hundred thousand dollars due to the rare light pink color and the special cutting. As the expert knew the diamond was made by them in the laboratory, he asked how much the cost for the diamond growth was. The thoughtless answer was about five thousand dollars. In fact, that was only the expense for the labor of one week (day and night) growth, and expenses for researches and equipment were out of consideration. On the other hand, the direct cost for raw materials might be even less than five thousand dollars. All that was needed was a small good natural diamond as seed only, methane together with hydrogen, and nitrogen gases as complements. The carbon atoms of methane accumulated on to the diamond seed in a microwave oven layer by layer, and thus the diamond grew up. After one week the diamond with light pink color was obtained. After laser cutting in the laboratory, they spent one hundred dollars for a master to make a brilliant cut. The diamond of a basic style was formed.

In 2002, C.-S. Yan, H.-K. Mao, R.J. Hemley et al. wrote the paper "*Very high growth rate chemical vapor deposition of single-crystal diamond*"[4] in *Proc. of National Academy of Science (PNAS)* of USA. At that time, the diamond crystals could already be produced by microwave plasma chemical vapor deposition (MPCVD) at growth rates up to 150 μm/h, which was 1–2 orders of magnitude higher than standard processes for making polycrystalline MPCVD diamond. The depositions were conducted in a 6-kW, 2.45-GHz MPCVD chamber with a volume of $\approx$5,000 cm^3, (Wavemat, model MPDR313EHP, Plymouth, MI) evacuated to a base pressure of $< 10^{-3}$ Torr (1 Torr = 133 Pa). The cavity and substrate stage had been redesigned to sustain stable and energetic plasma for high growth rate applications at low microwave power (1–2 kW) that also prevent plasma concentration at edges. The overall effect of the redesigned stage enhanced the growth rate by a factor of 10 and reduced the microwave power by a factor of 3 compared with Wavemat's original design. The deposition temperature was measured by a two-color infrared ($\approx$ 1.6 μm) ratio thermometer; the thermal emission was focused through a quartz window at an incident angle of 65° on the target, with a 2 mm diameter minimum target size. A PC automatically controlled the gas mass-flow, thermometer, and microwave-power level. The substrates were commercial $3.5\times3.5\times1.6$ mm^3 HPHT synthetic type Ib diamonds; the deposition surface was within 2° of the $\{100\}$ top surface. The substrates required polished, smooth surfaces that were cleaned ultrasonically with acetone and mechanically mounted on a molybdenum-substrate holder to ensure a uniform temperature. The deposited diamond was characterized by Raman and photoluminescence spectroscopes, electron paramagnetic resonance (EPR), and x-ray diffraction (XRD). Experimental variables including the reactant gas pressure and concentration, substrate temperature, and sample stage were varied to optimize diamond growth quality and increase $\langle 100\rangle$ growth rate. Their previous study showed that when the reactor pressure was increased from 60 to 200 Torr, the diamond growth rate increases $\approx$5-fold, while penetration twins decreased. A small amount of nitrogen added (1 %–5 % N_2/CH_4) to the reactive gas had been shown to have a beneficial effect of creating more available growth sites, enhancing the growth rate to three times and promoting {100} face growth. Using the published information on the optimal chemical conditions together with the redesigned stage, the growth-rate had been enhanced up to 2 orders of magnitude. The experiments were performed at a pressure of 160 Torr, 3 % N_2/CH_4, and a high percentage of methane (12 % CH_4/H_2). Under these conditions a seed produced with a 1-inch-diameter plasma, the growth morphology and color strongly depend on temperature, with the smooth, and dark brown at 1,000–1,100 °C, brown at 1,100–1,200 °C, and tinted yellow at 1,200–1,220 °C. Step flow-type with pyramid-like octahedral tinted yellow formed at 1,230–1,400 °C, spherical black diamond-like carbon was produced below 1,000 °C, and twinned or polycrystalline diamond resulted at over 1,300 °C. The temperature range for smooth, yellow-tint growth is narrow (1,200–1,220 °C) and can be missed easily.

Under these conditions, researchers at Carnegie can produce transparent MPCVD diamond that is as smooth as the HPHT diamond seed, enlarging both the thickness and area of the crystal. Fig. 9.4 shows a photograph of a $3.3\times3.5\times1.6$ mm^3 substrate seed and as-grown CVD $4.2\times4.2\times2.3$ mm^3 unpolished diamond with a 0.7 mm deposit on the seed in a 12 h run at a growth rate of 58 μm/h at 1,220 °C and 160 Torr [500 standard cm^3 min^{-1} (sccm) H_2, 60 sccm CH_4, and 1.8 sccm N_2] and microwave power of 1.6 kW. The growth morphology indicates that the $\langle 100\rangle$ side-growth rate is faster than that of the $\langle 111\rangle$ corner-growth rate. They had successfully produced 5-carat single crystal (dimensions, $7\times8\times5$ mm^3) after a ≈10-times-longer regrowth from a 0.3-mm-thick seed, but it was brown in color and had a crack on the {111} face. In 2005 they further obtained a 10 carat gem-quality single-crystal diamond.

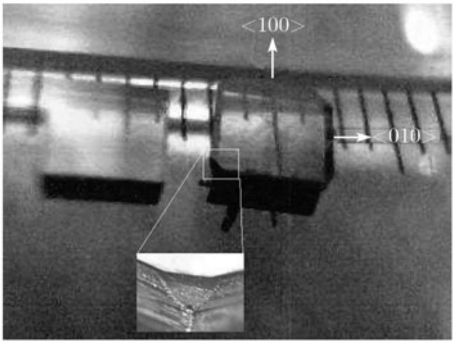


Fig. 9.4 Photograph of seed and as-grown unpolished CVD diamond and a magnification of CVD-diamond corner. The $\langle 100\rangle$ direction corresponds to the four sides of the diamond cube. Reprinted from C.S. Yan, Y.K. Vohra, H.K. Mao and R.J. Hemley [4] with permission from PNAS (Proc. the National Academy of Sciences)

For thermodynamic research, the conditions for the diamond phase growth are important, such as temperature, pressure and gas concentrations etc. Therefore, from the experimental results mentioned above it can be concluded that the temperature range 1,000–1,400 °C is suitable for the diamond growth, and the temperature range 1,200–1,220 °C is for the best diamond quality to be obtained. The concentration is 12 % methane in hydrogen, and the total pressure is 160 Torr, as shown in Table 9.1 (the nitrogen effect is not under consideration here).[4] Another important factor is the practical activation of microwave plasma, which will be discussed below.

Fig. 9.1 Experimental conditions for the low-pressure carat-size gem-quality diamond growth from the vapor phase

Pressure (Torr)	Temperature (°C)	Diamond quality	Concentration (CH_4/ H_2)
160	1200–1220	smooth, yellow-tint	12 %
160	1000–1400	single-crystal	12 %

For understanding the thermodynamic specialness of the activated low-pressure diamond growth, the first step was to do equilibrium phase diagram calculation using commercial software on the basis of Gibbs free energy minimization principle. The calculation software used by the author was EKVICALC.[6] Figure 9.5 shows a temperature-composition equilibrium phase diagram of carbon-hydrogen system (*T*–*X* equilibrium phase diagram of C–H system) at a fixed pressure of 160 Torr. Obviously, diamond is a metastable phase of carbon under low pressure, while graphite is a stable phase of carbon in this equilibrium phase diagram. This is because the phase region for the graphite deposition or graphite growth (in the right side of the solid line denoted by "gra") is bigger than the phase region for the diamond deposition or diamond growth (in the right side of the dashed line denoted by "dia"). The next step was adding Carnegie's experimental data for the carat-size single-crystal diamond growth (including temperature, pressure, concentration, but not including the nitrogen addition) into the equilibrium phase diagram of Fig. 9.5 for comparison. The reported region for the possible single-crystal diamond growth (i.e. 1000–1400 °C, 160 Torr and 12 % CH_4/H_2) was shown by a double arrow up and down. Meanwhile, the best conditions for the smooth, yellow-tint single-crystal diamond growth (i.e. 1200–1220 °C, 160 Torr and 12 % CH_4/H_2) were denoted by two open circles.

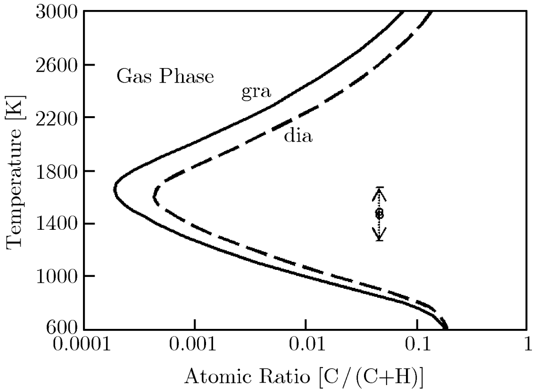


Fig. 9.5 C–H system *T*–*X* equilibrium phase diagram, pressure = 160 Torr (Carnegie's experimental data[4] are included)

Obviously, Carnegie's experimental data deeply dropped into the graphite and diamond phase regions of Fig. 9.5. It means that classical equilibrium thermodynamics can not provide a correct theoretical explanation, because the carat-size diamond growth system is a strongly activated nonequilibrium system but not an equilibrium system. Of course, both Yang's 2005 "nanothermodynamics" or also called "nanonucleate" model[7] and Hwang's 1996 "charged cluster" model[8] are impossible to provide correct explanations as well. Both of these models were based on stabilization of the diamond phase by the so-called "high pressure ($>$ 2 GPa)" due to the surface tension of the "nanonucleus" or "cluster" (radius $<$ 4 nm). However, now the millimeter-size or even centimeter-size diamond bulks are prepared by the millimeter plane-type epitaxial diamond growth. That is the same for Piekarczyk's 1998 "undersaturated vapor growth" model,[9, 10] because the ultrahigh concentration of 12 % methane in hydrogen is used in the high growth rate CVD process. The mistakes of these incorrect models mainly arise from the neglect of the important nonequilibrium activation due to the influences of classical thermodynamics.

According to Chapters 6–8 in this book, the corresponding nonequilibrium phase diagram, as shown in Fig. 9.6, has been quantitatively calculated on the basis of thermodynamic coupling model of nondissipative thermodynamics. From Fig. 9.6, the improvement of the microwave plasma chamber is the key point to provide stable high-efficiency low-power microwave plasma to ensure the high growth rate. After the improvement, the growth rate increased 10 times and the microwave power decreased to 1/3 of original power. That is, the activation was more effective due to the chamber improvement, so the concentration of superequilibrium atomic hydrogen was much higher than that in standard process. Therefore, the reaction rate ratio $\chi = 0.46$ was taken during the author's calculation of Fig. 9.6.

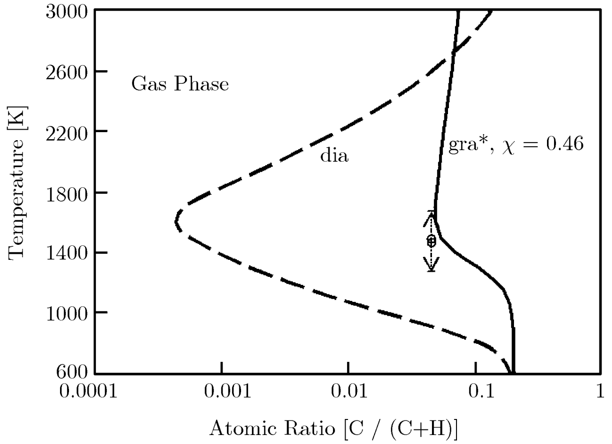


Fig. 9.6 Theoretical T–X nonequilibrium phase diagram for activated low-pressure diamond growth from C–H binary systems, p = 160 Torr, activated temperature = 3000 K and reaction rate ratio of reaction 2 to reaction 1, $\chi = 0.46$. (Carnegie's experimental data are included[4])

In our calculations of nonequilibrium phase diagrams for the ordinary activated low-pressure diamond growth from the vapor phase (published by us during 1970–2000 and described in Chapters 7 and 8), the value of reaction rate ratio $\chi = 0.28$ was taken (The plasma jet process was usually not included). The physical meaning of χ is very clear [refer to Eq. (6.12)–Eq. (6.17), or refer to Eq. (7.1)–Eq. (7.4)]. It is the reaction rate ratio between the association of superequilibrium atomic hydrogen and the transformation from graphite to diamond. It is closely connected with the practical concentration of superequilibrium atomic hydrogen, and is an important experimental parameter as the token of activation. It is also the value at one dimensional coordination for nonequilibrium condition in nonequilibrium phase diagrams. In Chapters 7 and 8 the taken value of $\chi = 0.28$ was mainly based on comparison with Matsumoto's experimental data of hot filament process, while the value of $\chi \geqslant 0.46$ was based on comparison with Carnegie's experimental data of the very fast growth rate process. The result seems to be reasonable, because it shows the effective improvement on the original MPCVD equipment of the Carnegie Institution to ensure the success of the"very high growth rate chemical vapor deposition of single-crystal diamond" process. Besides, during the calculation the activation temperature is needed. Lombardi et al. pointed out that the activation temperature of microwave plasma is about 3000 K, [11] so in the author's calculation the activation temperature = 3000 K was taken.

From Fig. 9.6 there is a diamond growth (or stable) region between the diamond phase line (dashed line denoted by "dia") and the activated graphite phase line (solid line denoted by "gra*, $\chi = 0.46$"), and Carnegie's experimental data for the high growth rate single-crystal diamond are just dropped in the diamond growth region. In comparison of Fig. 9.6 with other $\chi = 0.28$ C–H system T–X nonequilibrium phase diagrams in Chapter 7 (such as Fig. 7.9), it can be clearly found that the diamond growth region in $\chi = 0.46$ nonequilibrium phase diagram is much broader. From Fig. 7.8 the relationship between the diamond growth region and the reaction rate ratio χ in nonequilibrium phase diagram was clearly shown, and this will not be repeated here.

From the nonequilibrium phase diagram of Fig. 9.6 it can also be found that if $\chi \geqslant 0.46$ at the activation temperature = 3000 K, there seems to be no temperature limitation for the activated CVD diamond growth. In fact, inside the diamond bulk there is no influence of superequilibrium atomic hydrogen on phase stability of diamond, so diamond still subordinates the rules of equilibrium phase diagrams (i.e. diamond is still a metastable phase under low pressures inside the bulk). Therefore, at very high temperature, such as > 1800 K, the barrier of the spontaneous transformation from diamond to graphite under low pressure could still be over. Usually, high temperature is of benefit to single-crystal growth, but for the low-pressure diamond growth the temperature should not be too high.

To sum up, from the view point of modern thermodynamics, the success of Carnegie's microwave plasma CVD single-crystal diamond with very high growth rate and high-quality is closely connected with their high-efficiency reformation of microwave activation. The success is also due to the working pressure at 160 Torr, under which the carbon concentration in the vapor phase is much higher than that of ordinary microwave CVD process at $\approx$20 Torr, so the growth rate of diamond is much higher. Up to now, as the author knows, only our thermodynamic coupling model can provide a reasonable thermodynamic explanation. On the other hand, the low-pressure diamond process demonstrates the correctness and necessity of the development of modern thermodynamics and the new field of nondissipative thermodynamics.

9.2 Fluctuation of Equilibrium States and Stationary Nonequilibrium States

When an isolated system including a lot of molecules of ideal gas reaches its equilibrium state, entropy of the system must be a maximum and without any macroscopic change inside the system. However, inside practical systems the number of molecules is always limited. Molecules inside the system are continuously in movement, so "fluctuation" of macroscopic parameter at any location is possible and such a change should be under consideration.[12] Figure 9.7 shows the fluctuation in an isolated single-gas equilibrium system. An arrow in the figure illustrates a fluctuation due to the possible departure of molecules from their equilibrium location. When the molecule with arrow moves toward the left side, the concentration difference forms inside the system, the system is ordered, and the total entropy of the system decreases. According to the entropy increase principle, the mathematical expression of the second law of thermodynamics, entropy of the isolated system is always spontaneously toward the direction of entropy increase. Therefore, the system spontaneously comes back to its original uniformity, i.e. the system comes back into its most disordered state, which is homogeneous equilibrium state. When the molecule with arrow moves toward the right side, the concentration difference also forms inside the system, the system is ordered, and the total entropy of the system decreases. According to the entropy increase principle, the mathematical expression of the second law of thermodynamics, the total entropy of the isolated system is always spontaneously toward the direction of entropy increase. Therefore, the system also spontaneously comes back to its original uniformity, i.e. the system comes back into its most disordered state, which is homogeneous equilibrium state. In stability theory of equilibrium, such an equilibrium is stable.

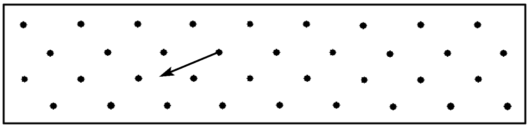


Fig. 9.7 Illustration of fluctuation in an isolated single-gas equilibrium system

Figure 9.8 shows fluctuation in an isolated binary-gas-mixture equilibrium system. The fluctuation is basically the same as that in Fig. 9.7. In spite of few molecules depart from their equilibrium locations to the left or to the right, concentration difference forms inside the system, the system is ordered, and the total entropy of the system decreases. According to the entropy increase principle, the mathematical expression of the second law of

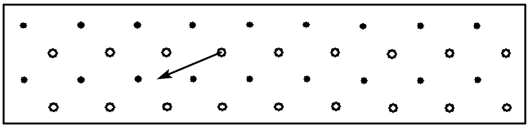


Fig. 9.8 Illustration of fluctuation in an isolated binary-gas-mixture equilibrium system

thermodynamics, the total entropy of the isolated system is always spontaneously toward the direction of entropy increase. Therefore, the system also spontaneously comes back to its original uniformity, i.e. the system comes back into its original equilibrium state. Such an equilibrium state is also stable. Fluctuation in both the single-gas system and the binary-gas-mixture system is very similar.

Figure 9.9 shows fluctuation in a stationary nonequilibrium thermodiffusion binary-gas-mixture system. If at a constant temperature (i.e. $T_1 = T_2$) there is no heat flow through the system, the closed system will be a binary uniform gas mixture in equilibrium. If $T_1 > T_2$, after a certain time a stable heat flow goes through the system, and a concentration gradient appears inside the binary gas mixture. Usually, the lighter molecules will concentrate to the high temperature side, while molecules of another gas will concentrate to the low temperature side. Finally, a stationary nonequilibrium state, as shown in Fig. 9.9, will be obtained. The arrow in Fig. 9.9 shows the movement of an individual light molecule departing from its location of the stationary nonequilibrium state. If it moves toward the left side, concentration difference will be bigger inside the system, and the system is more ordered, but if it moves toward the right side, concentration difference will be smaller inside the system, and the system is more disordered. The results due to such a diffusion movement are quite different from each other in the stationary nonequilibrium state of Fig. 9.9 and in the equilibrium state of Fig. 9.8. That shows the different effect of diffusion. In the stationary nonequilibrium state of Fig. 9.9 the state function is not in extremum, or we may say the equilibrium extremum principle is not suitable for such a stationary nonequilibrium state.

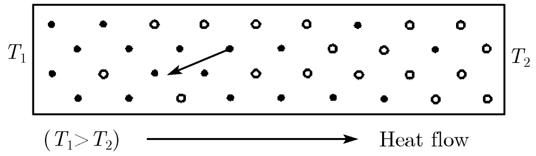


Fig. 9.9 Illustration of fluctuation in a stationary nonequilibrium thermodiffusion binary-gas-mixture system

9.3 Some Discussions on Classification of Thermodynamics

The classification of thermodynamics attracted attention perhaps in the second half of the 20th century. In particular, after Onsager and Prigogine won their Nobel Prizes, scientists knew their work was obviously different from thermodynamics of the 19th century. Thus thermodynamics of the 19th century (maybe up to the beginning of the 20th century) was called classical thermodynamics. However, there were no clear definitions for both classical thermodynamics and nonclassical thermodynamics. That is perhaps the reason why there seemed to be only “classical thermodynamics” but no “modern thermodynamics”. Such a strange situation for thermodynamics lasted for a long time.

Without any question, systems in Onsager’s and Prigogine’s researches were nonequilibrium systems, while systems in researches of classical thermodynamics were mainly equilibrium systems. Therefore, it was natural to form a classical classification of thermodynamics, as shown in Fig. 4.4. That is, the thermodynamics discipline had been divided into two main parts: equilibrium thermodynamics and nonequilibrium thermodynamics. The main shortcoming of this was that it was difficult to find a suitable location and term for the inequality part of classical thermodynamics. For instance, the examples of gas expansion to vacuum, mixture of two gases and supercooling solidification in Section 3.6 of this book should also be included in nonequilibrium thermodynamics. However, these examples are quite different from those in Onsager's and Prigogine's researches.

In order to distinguish Onsager's and Prigogine's work from thermodynamics of the 19th century, a pair of "thermodynamics for equilibrium states" and "thermodynamics for nonequilibrium states" had been proposed. However, such a pair of terms is very close to another generally used pair of "equilibrium thermodynamics" and "nonequilibrium thermodynamics" in Chinese. These two pairs of terms are often confused with each other. For instance, in a book it was written that [13]

> Based on different states of thermodynamic systems can be divided into systems in equilibrium states and systems in nonequilibrium states, so equilibrium thermodynamics and nonequilibrium thermodynamics are formed.

Some thermodynamicists also proposed a pair of "thermodynamics for reversible processes" and "thermodynamics for irreversible processes". Such a pair of terms was used for distinguishing the character of Onsager's and Prigogine's researches because thermodiffusion and chemical oscillations are both concerned with ongoing nonequilibrium states or irreversible processes taking place. Meanwhile, the shortcomings and limitation of classical thermodynamics had also been drawn-out a little. In classical thermodynamics the equality of the second law is surely for equilibrium state of system, while the inequality of the second law in classical thermodynamics (such as for gas expansion to vacuum, mixture of two gases and supercooling solidification) is also only concerned with their initial equilibrium states and their final equilibrium states. Classical thermodynamics makes research neither on the nonequilibrium state itself nor on irreversible process itself. As mentioned-above, that was a key restriction of classical thermodynamics, but it is difficult for us to make a further division in "thermodynamics for equilibrium states". That is because the terms of "equilibrium thermodynamics in thermodynamics for equilibrium states" and "nonequilibrium thermodynamics in thermodynamics for equilibrium states" for equality and inequality of classical thermodynamics, respectively, would be very bad and should not be taken.

To prevent confusion of terms the author has made prior efforts to use generally accepted terms, such as classical thermodynamics, modern thermodynamics, equilibrium thermodynamics, nonequilibrium thermodynamics etc. However, the recognition of Fig. 4.9 A basic and complete modern classification of thermodynamics had also been through a long way for eight years (2000–2008).

From 2000 to the beginning of 2002 the author found a new field of nonequilibrium nondissipative thermodynamics, and then added the new field into the equality part of the second law of thermodynamics on the basis of the original traditional classification of thermodynamics (i.e. Fig. 4.5 Traditional classification of thermodynamics). The result after the addition is shown in Fig. 9.10.

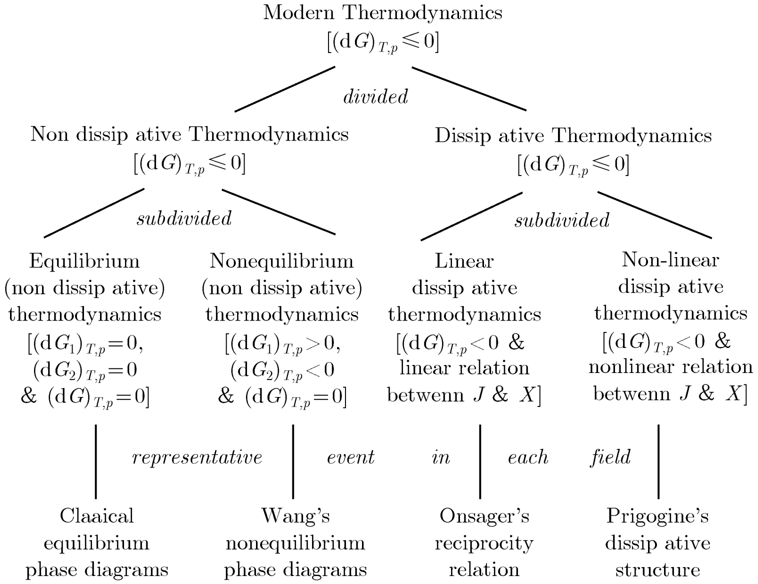


Fig. 9.10 An classification of thermodynamics accepted in author's book published by Springer in 2002

At that time such a classification of thermodynamics emphasized the new field of nonequilibrium nondissipative thermodynamics belonging to the equality part of mathematical expression of the second law of thermodynamics, so nonequilibrium phase diagrams could be quantitatively calculated in the new field. That was successful, but there were also shortcomings in that classification. First, that classification was not a complete one, and nonequilibrium thermodynamics part of classical thermodynamics had been neglected. Second, there was no clear distinction between classical thermodynamics and modern thermodynamics, and there was no clear overview of thermodynamics. As that classification, as shown in Fig. 9.10, had been used in the author's 2002 Springer book, "Nonequilibrium Nondissipative Thermodynamics", the author sincerely hopes to remove any possible negative influences resulting from this.

From mid-2002 to mid-2007 the author had basically used the complete classification of thermodynamics, as shown in Fig. 9.11.[12, 15, 16] However, the criteria for both thermodynamics discipline and classical thermodynamics were the same. That was not suitable. Up to 2007, in the *Proc. of Chinese National Symposium on Thermodynamics and Statistical Physics* the classification of Fig. 9.11 was still used. On 6 August 2007 at a plenary lecture of this National Symposium given by the author, further improvement was shown, and the complete classification of thermodynamics, as shown in Fig. 9.12, had been practically used. The mathematical expressions of the whole of thermodynamics and classical thermodynamics were $[\mathrm{d_i}S \geqslant 0]$ and $[\mathrm{d_i}S_\mathrm{p} \geqslant 0, \mathrm{d_i}S \geqslant 0]$, respectively. Here, $\mathrm{d_i}S$ and $\mathrm{d_i}S_\mathrm{p}$ are entropy productions of system and internal process, respectively.

There was also a notice that in classical thermodynamics entropy production of process(es) $\mathrm{d_i}S_\mathrm{p}$ and entropy production of system $\mathrm{d_i}S$ must be of the same value and the same sign. Therefore, at the end of 2007 the mathematical expression of classical thermodynamics had further been improved into $[\mathrm{d_i}S_\mathrm{p} = \mathrm{d_i}S \geqslant 0]$, in which the condition that entropy production of process(es) $\mathrm{d_i}S_\mathrm{p}$ and entropy production of system $\mathrm{d_i}S$ must be of the same value and the same sign had already been included, as shown in Fig. 1.1 or Fig. 4.9.

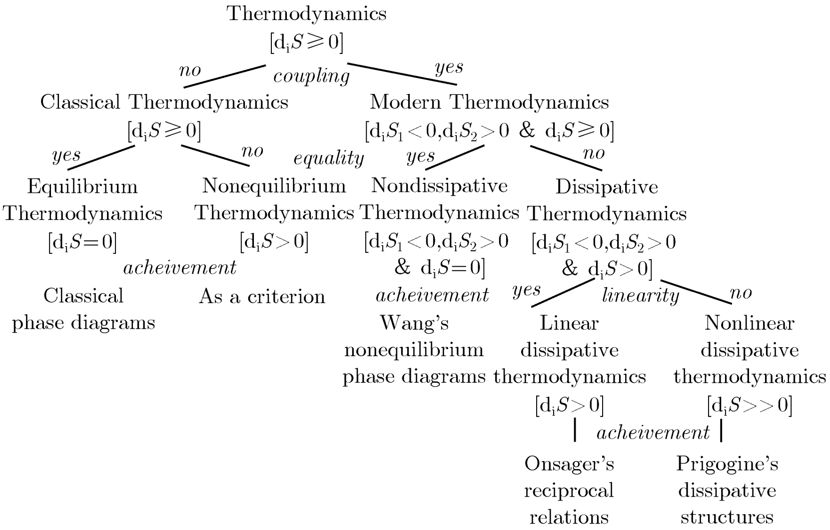


Fig. 9.11 The complete classification of thermodynamics used by the author during mid-2002–mid-2007. [12, 15] It was not suitable to use the same mathematical expressions for thermodynamics and classical thermodynamics.

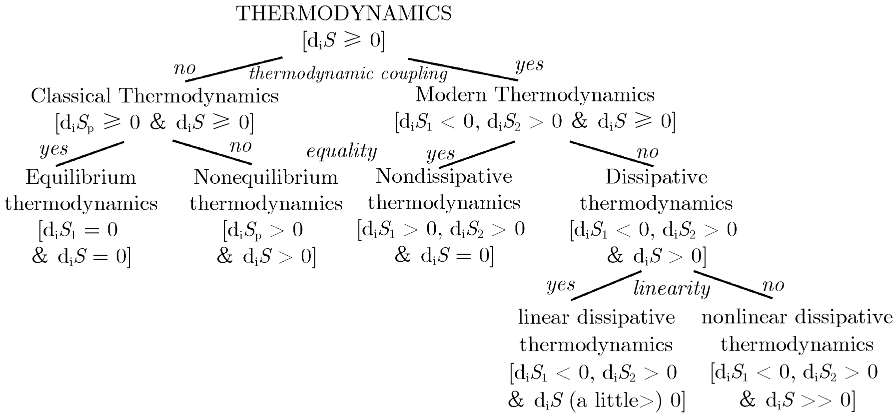


Fig. 9.12 The complete classification of thermodynamics used by the author as of 2007. [12, 15] Entropy productions of both process $\mathrm{d_i}S_\mathrm{p}$ and system $\mathrm{d_i}S$ are of the same value and the same sign in classical thermodynamics.

If someone is not familiar with, or is doubtful of the form of entropy production, please convert Fig. 4.9 into the form of Gibbs free energy under isothermal isobaric conditions by using Eq. (1.4), as shown in Fig. 9.13, which may be more easy to be understood.

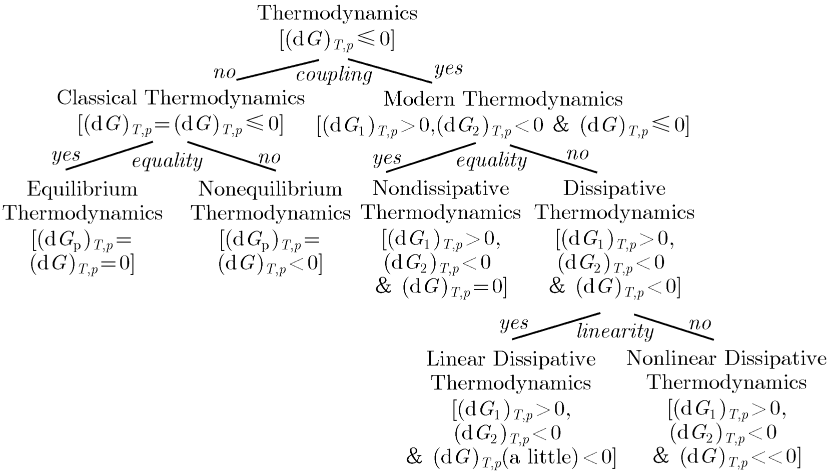


Fig. 9.13 A basic and complete classification of thermodynamics under isothermal isobaric conditions. $(\mathrm{d}G_\mathrm{p})_{T,p}$, $(\mathrm{d}G_1)_{T,p}$ and $(\mathrm{d}G_2)_{T,p}$ are Gibbs free energy changes of processes, while $(\mathrm{d}G)_{T,p}$ is Gibbs free energy change of system.

9.4 What is "Thermodynamics" and What is "the Second Law of Thermodynamics"

In the last paragraph of the first section in Chapter 1 "1.1 Challenges to the Second Law of Thermodynamics Coming from Two Sides", it was said that

> To sum up, what is "thermodynamics" and what is "the second law of thermodynamics" become the key points of puzzlement.

This had been explained through out the book, and so at this point in the text it should now be clear.

At the beginning of Chapter 5 it was said that

> Thermodynamics is a discipline of science for research on development and evolution of macroscopic systems.

The definition or written statement of thermodynamics in Chapter 1 is that

> Thermodynamics is a part of science doing research on energy and energy transformation, and its research objects are the change, development and evolution of macroscopic systems consisting of a large number of particles.

The first half of the sentence was cited from "*Thermodynamics*" written by J.P. Holman,[17] while the second half of the sentence was cited from many textbooks on thermodynamics. The detailed contents of this book could be summarized into "Modern Thermodynamics — Based on the Extended Carnot Theorem". Here, "modern thermodynamics" is an exception in this book with a general meaning of the modernized thermodynamics discipline in the current 21st century.

Thus, the question what is “the second law of thermodynamics” may still remain. First, scientists (or better to say, thermodynamicists, maybe including the author) usually call both “the written statement of the second law of thermodynamics” and “the mathematical expression of the second law of thermodynamics” simply as “the second law of thermodynamics”. In fact, even just for “the written statement of the second law of thermodynamics” there are different descriptions and different understandings. In this book only the written statement in Chapter 1, i.e. “the second kind of perpetual motion machines are impossible to obtain” was emphasized. Another more active statement for fundamental laws of thermodynamics can also be accepted, i.e. “Except the first kind and the second kind of perpetual motion machines any kind of macroscopic energy transformation or any macroscopic state change of systems composed of a large number of particles is possible to realize.” Of course, examples or detailed descriptions about what the second kind of perpetual motion machines are should be attached, The examples or detailed descriptions become the key to understand the second law of thermodynamics. However, in most of textbooks of thermodynamics, the examples of the second kind of perpetual motion machines are only concerned with the characteristic direction of single spontaneous processes, such as the so-called Clausius’ statement, i.e. “heat cannot of itself, without the intervention of any external agency, pass from a colder to a hotter body”, or the so-called Kelvin’s statement, i. e. “it is impossible to obtain work by cooling a body below the lowest temperature of the system”, and so on. Therefore, the most important Clausius’ statement on “compensation” or now called “thermodynamic coupling”, as mentioned in Chapter 1, should be added into the written statement of the second law of thermodynamics as follows.

The second law of thermodynamics asserts that all spontaneous processes may take place (in a positive direction) by themselves without compensation (i. e. for simple uncoupling systems), such as heat conduction from the hotter body to the colder body; and any nonspontaneous process (in the negative direction) can only take place in such a manner as to be compensated by simultaneous occurring spontaneous process (es) (i. e. for complex coupling systems).

9.5 Thermodynamic Weakness of Physicists — Complex Systems

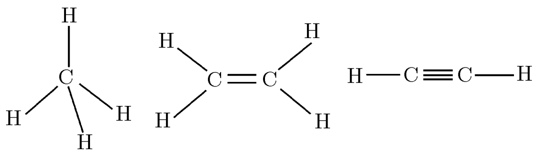
Scientific systematization has played an important promotion role in the development of science. Natural sciences are divided into the different disciplines of mathematics, physics, chemistry, biology etc. Correspondingly, there are different departments or colleges of mathematics, physics, chemistry, biology etc. in higher-education organizations, and after training students become mathematicians, physicists, chemists, biologists et al. In practical situations, they are often isolated from each other based on scientific systematization and have a lack of knowledge about neighboring disciplines. However, the development of nature and the objective world is never isolated on the basis of disciplines of science. In particular, for complex systems, except from a few example, such as thermodiffusion, nearly all complex systems include chemical reaction. Complex systems concerned with chemical reaction, especially in the life sciences were often regarded by most physicists as a no-no. However, the situation is slowly changing. In our Fudan University, undergraduate students learn basic knowledge on so different disciplines of science in the first year. Some famous physicists have covered the discipline borderline and directly come into the life sciences, or called "bio-X" (biochemistry, biophysics, etc.). However, due to historical influences, thermodynamics for complex systems are still the weakness of most physicists. Such situations may not only be a concern for the forum of physicists, but also for those in chemistry, biology and other disciplines. Development in modern science including modern thermodynamics is often concerned with several disciplines of science.

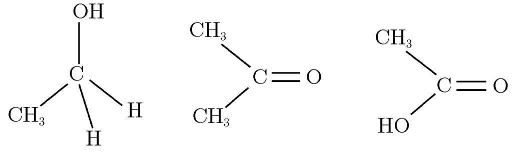
The example of fuel battery mentioned above could still be discussed. The complete transformation from chemical energy of methane-oxygen (CH_4–O_2) combustion to electrical energy must be of the highest conversion efficiency of energy. It is not necessary to pursue "reversibility". Therefore, the extended Carnot theorem means that any nondissipative process or system is of the highest conversion efficiency of energy. Such a simple example has already included knowledge from different disciplines. For practical calculations of an example of nonequilibrium phase diagrams for the low-pressure diamond growth, only a whole coupled reaction and two single-element reactions (1. carbon element: low-pressure nonspontaneous transformation from graphite to diamond and 2. hydrogen element: spontaneous association of superequilibrium atomic hydrogen) are concerned. However, the low-pressure diamond growth from the vapor phase had puzzled some physicists and other scientists for a long time (some decades since the beginning of 1970s).

The application of dissipation decrease theorem concerns more disciplines of science. In a single-gas system, thermodiffusion, as shown in Fig. 5.1, will not occur. The realization of dissipation decrease theorem should be based on a certain internal (perhaps concerned with microscopic) condition. In the low-pressure CVD diamond growth system, just due to the coexistence of (1) nonspontaneous reaction from graphite to diamond, and (2) spontaneous association of superequilibrium atomic hydrogen, the isothermal isobaric Gibbs free energy released from reaction (2) could be accepted by reaction (1) for the low-pressure diamond growth to realize dissipation decrease theorem. In (isothermal isobaric) B-Z reactions, the appearance of spiral reaction for Gibbs free energy decrease is also due to the complexity of B-Z reaction systems, in which a lot of possible simultaneous reactions could be chosen for transformation of Gibbs free energy from a spontaneous reaction to another nonspontaneous reaction. In a simple reaction system the curves of Gibbs free energy decrease may simply go down to the end of reactions. Therefore, classical thermodynamicists had been much puzzled by chemical oscillations. The complexity of living bodies provides excellent conditions to realize dissipation decrease theorem. The energy produced by metabolic oxidation process in living bodies may be transferred and stored into a form of proton (H^+) concentration gradient across the inner membrane of mitochondria. When protons H^+ spontaneously transfer from the high concentration side (outside) to the low concentration side (inside) of the inner mitochondrial membrane through a channel of ATPase molecule, the nonspontaneous synthesis of ATP (see Fig. 4.1–Fig. 4.3) could be performed. Complex structures of mitochondria are physical bases to complete such a complex energy transformation and strongly improve the energy efficiency. Without certain fundamental knowledge of chemistry and biochemistry physicists have difficult to treat such complex thermodynamic projects.

From the viewpoint of thermodynamics, together with necessary fundamental knowledge of the chemical periodic table of the elements, it can be predicted that living bodies (including the present living bodies on the earth and possible living bodies on other planets) must have a certain commonness — constructed mainly by compounds of carbon atoms together with atoms of hydrogen, oxygen, nitrogen, etc. In fact, different living groups on the earth have already been found. For instance, pyrodictium (P. brockii) is a genus of submarine hyperthermophilic archaebacterium, which was originally isolated from the solfatara fields off the coast of Volcano, Italy. The optimal growth conditions for P. brockii are temperature ranges, approximately 80–105 °C; pH ranges, 5–7; salt concentrations, 1.2 %–12 %; and through a form of metabolism known as hydrogen-sulfur autotrophy, which is characterized by the oxidation of H_2 by S° (element sulfur) to produce ATP and H_2S. P. brockii is a kind of strictly anaerobic bacteria, inefficient after few minutes explosion in oxygen.[18] Another example of marine sulphate-reducing bacteria was reported by *Nature* on 18 October 2007.[19] Propane or n-butane could be used as the sole growth substrate under oxygen-lack conditions. Marine sulphate-reducing bacteria with such capacity was enriched from marine hydrocarbon seep areas, such as Mexico bay and California bay. However, all of these living groups are composed of protein, DNA, RNA and other similar complex components, because complex components with high microscopic selectivity are necessary for coupling and energy utilization.

Carbon is located in the fourth group at the middle of the second period, so usually can form four σ covalent single-bonds toward peaks of a tetrahedron (i.e. four sp^3 covalent single-bonds); three σ covalent single-bonds toward tips of a plane triangle (i.e. three sp^2 covalent single-bonds) and one π bond vertical to the plane triangle for double-bond formation; or two linear profile covalent single-bonds (i.e. two sp covalent single-bonds) and other two vertical to each other π bonds for triple-bond formation, as shown in Fig. 9.14 (a)–(c). Their representations are: (a) methane, CH_4; (b) ethylene, C_2H_4; and (c) acetylene, C_2H_2. If one or more hydrogen atoms are substituted by carbon atom(s), oxygen atom(s), nitrogen atom(s) or their group(s), then the


(a) methane (b) ethene (c) acetylene


(d) alcohol (e) acetone (f) acetic acid

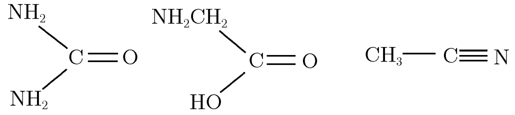
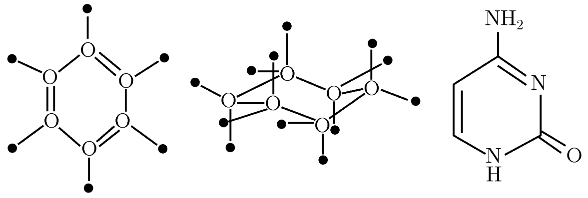
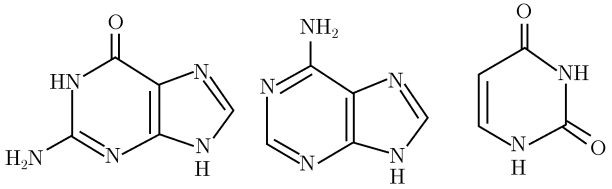

(g) urea (h) glycin (i) acetonitrile

Fig. 9.14 Different bonds of carbon atoms and some simple C-, O-(or N-) containing organic compounds following compounds can be obtained, such as (d) alcohol, C_2H_5OH; (e) acetone, $(CH_3)_2CO$; (f) acetic acid, CH_3COOH; (g) urea, $(NH_2)_2CO$; (h) glycin, NH_2CH_2COOH; and (i) acetonitrile, CH_3CN etc. Most of them are organic compounds. Urea is an inorganic compound (carbon-hydrogen bond is not included), but a product of metabolism of life. Among them (h) glycin is a representation of α-amino acids (i.e. H_2N-CHR-COOH). R radical is located close to carboxylic radical at the called α position. Different α-amino acids are due to different R radicals. Different proteins of living bodies on the earth are mainly composed of 20 basic α-amino acids through peptide chain (-HN-CHR-CO-) by polymerization.

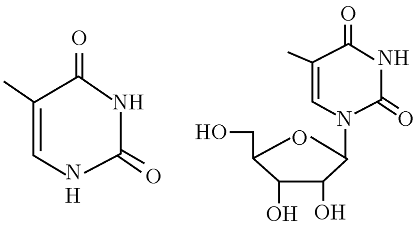
The bond character of carbon atoms makes it possible for carbon atoms to form various ring compounds, refer to Fig. 9.15. Figure 9.15(a) shows aromatic molecule of benzene (C_6H_6), while Fig. 9.15(b) shows aliphatic cyclohexane (C_6H_{12}). Here, the open circle represents carbon atom, while the small solid dot represents hydrogen atom. In the aspect of structures, benzene is similar to graphite, and cyclohexane is similar to diamond. Just due to the difference in their structures, atomic hydrogen could play an important role in coupling, as discussed in Chapter 6. Besides, the structure type of benzene is very important because there are alternating single-double bonds (in some aromatic molecules there are alternating isolated pair p electrons and double bonds). The valence electron



(a) C_6H_6 (b) C_6H_{12} (c) Cytosine



(d) Guanine (e) Adenine (f) Uracil



(g) Thymine (h) β-thymine-D-nucleoside

Fig. 9.15 (a) Benzene, (b) cyclohexane, (c)–(g). bases in DNA or RNA (C, G, A, U, T) , and (g) nucleoside distributions are practically and basically uniform along these alternating single-double bonds. That is, the moving ability of π electrons in double bonds or in isolated pairs of p electrons is very strong. These alternating single-double bonds are called conjugative bonds or called big π bonds, which may play an electric conductive role to transfer information in living bodies. Therefore, these conjugative bonds or called big π bonds are very important to perform life functions. From Fig. 4.1 the electron transport chain in the inner mitochondrial membrane and mechanisms of oxidative phosphorylation, one can find the important role of the electron transport. Otherwise, thermodynamic coupling of ATP biosynthesis due to protons passing through enzyme on the inner membrane would not be realized.

Thermodynamics is a macroscopic science. In principle, microscopic processes (including kinetic mechanisms) may be out of consideration. However, if the microscopic kinetic mechanisms had already previously been understood well, then the kinetic mechanisms may be of great help. An understanding of the nature of B–Z reactions, spiral reactions, was also due to the help of the Field-Körös-Noyes mechanism (i.e. FKN mechanism) of B–Z reactions. Macroscopic phenomenon of complex systems is more closely connected with microscopic mechanisms. Thermodynamics and kinetics are helpful to each other or compansate for each other. Even for different disciplines (physics, chemistry, and biology etc.) they are closely related with each other. In general, we should never absolutely isolate them.

Figure 9.15 (c) – Fig. 9.15 (g) are five bases in the most important hereditary materials, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). They are: Cytosine (symbol C, $C_4H_5N_3O$), Guanine (symbol G, $C_5H_5N_5O$), Adenine (symbol A, $C_5H_5N_5$), Uracil (symbol U, $C_4H_4N_2O_2$) and Thymine (symbol T, $C_5H_6N_2O_2$), respectively. Bases in RNA are C, G, A, and U; while bases in DNA are C, G, A, and T. Note that, DNA is the carrier of gene. The sequence of bases (for example, CAG) is the genetic code. All of these bases are aromatic heterocyclic organic compounds including big π bonds, in which electron activity is very strong. All of these radicals may be some kind of molecular device units, which we haven't understood well. Base and ribose may connect to each other to form the corresponding nucleoside, such as Fig. 9.15 (h) shows β-thymine-D-nucleoside. Positions in the base of nucleoside are numbered as $1, 2, 3, \cdots$, while in ribose only carbon positions (bold bonds in the front) are numbered clockwise as 1', 2', 3', 4', 5'; and β- and D- are concerned with configurations of carbon atoms 1' and 4'. In structure figures of complex organic molecules, carbon atoms together with the connected hydrogen atoms are usually not necessary to draw out. In ribonucleic acid (RNA) molecules, one composition, ribose, is D-ribose, which is a monosaccharide with molecular formula $C_4H_9O_4CHO$. Nitrogen atoms in bases (A, G, C and U) of RNA molecules are connected with carbon atoms of D-ribose molecules to form the corresponding nucleoside. 3'-hydroxyl in nucleoside molecules and/or (for the head and the tail of the molecular chain "or" should be used instead of "and") 5'-hydroxyl are esterified by phosphoric acid, and each phosphate group is connected with nucleoside to form a phosphate chain or strands, i.e. RNA. In molecules of deoxyribonucleic acid (DNA), one composition, ribose, is 2'-deoxy-D-ribose, or called deoxyribose. Its molecular formula is $C_4H_9O_3CHO$. Nitrogen atoms in bases (A, G, C and T) of DNA molecules are connected with carbon atoms of deoxyribose molecules to form the corresponding nucleoside. 3'-hydroxyl in nucleoside molecules and/or (for the head and the tail of the molecular chain “or” should be used instead of “and”) 5’-hydroxyl are esterified by phosphoric acid, and each phosphate group is connected with nucleoside to form a phosphate chain or strand, i.e. a half of DNA. DNA is of double helix structure, as shown in the upper half of Fig. 9.16. Hydrogen bonds between the base pairing connect two strands wound around each other. The rules of base pairing are: A connects with T by two hydrogen bonds, and C connects with G by three hydrogen bonds. The shape and size of the A–T base pairing are near the same as that of C–G base pairing.

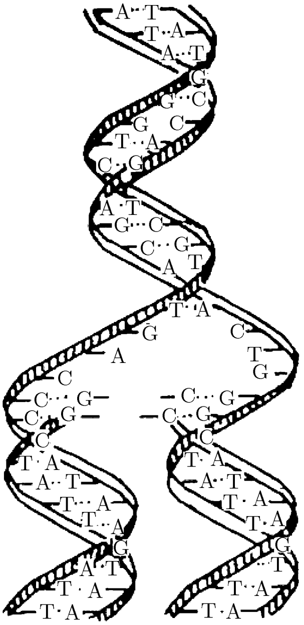


Fig. 9.16 The double helix structure of DNA and replication of DNA

During the stage of cell division, both helicase and polymerase may coexist. Four deoxynucleotides (corresponding to A, G, C and T) are stored as materials, and ATP provides energy. The DNA double-helix may be unwound into replication fork by helicase at the middle, as shown in Fig. 9.16. Meanwhile, on the two single strands two complementary strands are synthesized individually. The original double helix DNA becomes two DNA double helix DNA, as the bifurcate point moves up. This is replication of DNA. The famous technique of PCR (Polymerase chain reaction) for judicatory identification is replication of DNA under manual conditions (different from conditions in cell). From the viewpoint of thermodynamics, the original four deoxynucleotides (corresponding to four bases) are in disorder. During replication, “selforganization” or ordering process, the energy providing, such as the transformation from ATP to ADP, are needed. The nonspontaneous replication of DNA and the spontaneous transformation from ATP to ADP take place simultaneously. That is thermodynamic coupling. Therefore, the replication of DNA, which is closely connected with important living phenomena (inheritance, aberrance, propagation etc.), is only a nonspontaneous ordering process of thermodynamic coupling. There is nothing violating the second law of thermodynamics.

The hereditary material in a cell is coded in the sequence of four bases of DNA. There are normally 46 strands of DNA called chromosomes in human cells. Specific regions, called genes, on each chromosome contain the hereditary information which distinguishes individuals from each other. The sequence of bases is the first-order structure, the double helix is the second-order structure, and there are the higher-order structures (such as folding) of DNA. Another important substantial base of life is protein, which is composed of 20 basic amino acids. Similarly, the sequence of amino acid is the first-order structure, the helix is the second-order structure, and there are also the higher-order folding structures of proteins. Such complex structures of DNA and proteins provide the necessary substantial bases to realize thermodynamic coupling, dissipation decrease theorem, so energy transformation in living bodies is usually of very high efficiency. Just as mentioned in Chapter 4, the chemiosmotic theory, or called the chemiosmotic coupling theory, is really a theory of thermodynamic coupling.[20] Up to now in the life sciences the double arrows have generally, qualitatively and very successfully been used for thermodynamic coupling, as shown in Fig. 9.17 Shuttle reactions in mitochondria, and in Fig. 4.1–Fig. 4.3.[21, 22] However, all of these examples could not be explained in classical thermodynamics.

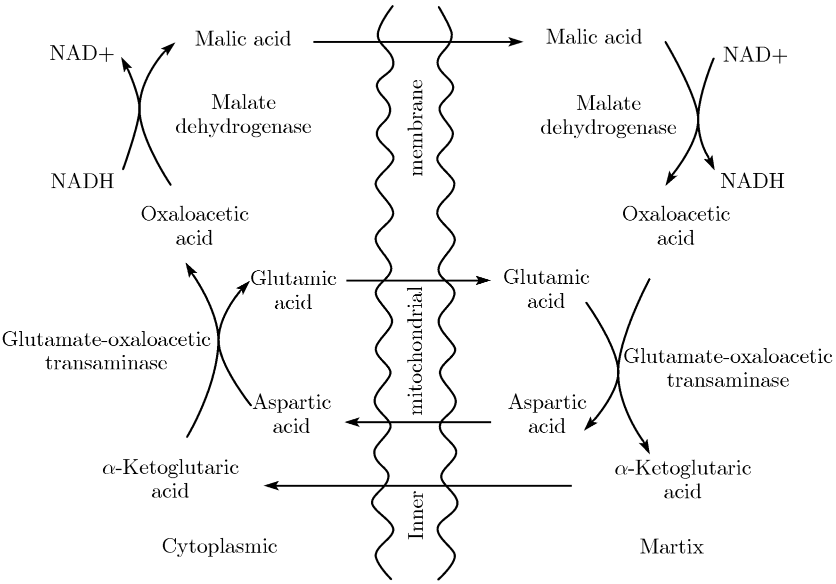


Fig. 9.17 Shuttle reactions in mitochondria

It is pity that some classical thermodynamicists, including some physicists (and scientists in other disciplines) know too little about thermodynamic coupling (or called reaction coupling) widely in the life sciences, and also too little about the development of thermodynamics for complex systems. "Reaction" in reaction coupling (i.e. thermodynamic coupling including chemical reaction) becomes the key point of their weakness. Up to now thermodynamic coupling of the low-pressure diamond is the simplest practical example of reaction coupling (the mechanism is clear at the atom level – refer to Fig. 6.25– including only two kinds of elements: carbon and hydrogen). However, it may still be regarded by some physicists as a "no-no". Now it is the time for us to improve the unsuitable situation between theoretical and practical demands, to change the incorrect expression of $[(\Delta G_1^0)_{T,p} > 0,\ (\Delta G_2^0)_{T,p} < 0\&\ (\Delta G^0)_{T,p} \leqslant 0]$ into the correct expression of $[(\Delta G_1)_{T,p} > 0,\ (\Delta G_2)_{T,p} < 0\&\ (\Delta G)_{T,p} \leqslant 0]$ for thermodynamic coupling (refer to Chapter 4), and to make contributions to quantitative researches on thermodynamics in biology. For a few thermodynamicists, who are already working in the life sciences, the following situation could be regarded as a caution. In a plenary lecture about "thermodynamics in biology" at *the 2000 International Conference on Chemical Thermodynamics* (ICCT2000), thermodynamics data and equations of ΔG^0 were confused with those of ΔG (this is also a general problem in books on thermodynamics in biology). After his presentation the speaker had difficulty answering the question about "Why two different conceptions (ΔG^0 and ΔG) could be confused?" That had been one of the hot points at that conference. The duty of thermodynamicists now, including some physicists and scientists of other disciplines should be to do more on thermodynamics in biology and biochemistry.

9.6 About "Nonequilibrium Thermodynamics of Small Systems"

As early as 1931, D. Burk had already suitably talked about similar projects. He said: [23]

> The problem of the general applicability of the second law of thermodynamics to biological systems has of late years received attention of substantial character. The two following different but hardly conflicting views summarize the present position. Lewis and Randall state, 'The second law of thermodynamics is a principle which has never failed to satisfy the severest test of experiment.' Donnan states, as a result of probability calculations, 'It seems, therefore, that there exist biological systems of such minute dimensions that the laws of classical thermodynamics are no longer applicable to them.' It is the purpose of the present paper to offer some rather exact support for the first of these statements, without, however, detracting an any way from the suggestiveness of the second.

The title of this section, "*The nonequilibrium thermodynamics of small systems*", was cited from the title of a paper in July, 2005, published in *Physics Today*. [24] From the title, the point of view seems to be unlike the two points of view mentioned above by Burk. In fact, all of them are "different but hardly conflicting views". There were only two sentences in the abstract of this paper. The first of the two sentences here said:

> The interactions of tiny objects with their environments are dominated by thermal fluctuations.

However, the fluctuation is just the deviation of thermodynamic predictions. As said in this paper:

> As a system's dimensions decrease, fluctuations away from equilibrium begin to dominate its behavior. In particular, in a nonequilibrium small system, thermal fluctuations can lead to observable and significant deviations from the system's average behavior. Therefore, such systems are not well described by classical thermodynamics.

Note that, fluctuation in both nondissipative thermodynamics and dissipative thermodynamics of modern thermodynamics might also result in obvious deviation of average. However, the statistical average will still subordinate thermodynamic laws.

The second sentence of the two sentences abstract said:

> Guided by theory (of modern fluctuation) and assisted by new micromanipulation tools, scientists have begun to study such interactions in detail.

It means that the research under the title of "*The nonequilibrium thermodynamics of small systems*" seemed to be mainly on microscopic kinetic mechanism and fluctuation, but conclusions may still be related with macroscopic thermodynamic performance. That may be the reason to put the term of "nonequilibrium thermodynamics" into the title. Section 4.7 Chemiosmotic Coupling Theory for ATP Biosynthesis and Fig. 4.1–Fig. 4.3 in Chapter 4; Section 6.6 Mechanism of Thermodynamic Coupling in Low-Pressure Diamond Growth and Fig. 6.25 in Chapter 6; Section 9.2 Fluctuation of Equilibrium States and Stationary Nonequilibrium States and Fig. 9.16 in the last section are related with "The nonequilibrium thermodynamics of small systems". The double helix structure of DNA and replication of DNA, and Fig. 9.17 Shuttle Reactions in Mitochondria in this Chapter may all belong to "The nonequilibrium thermodynamics of small systems". Conclusions may all be concerned with energy transformation, but the research on molecular machine, especially on molecular motor guided by modern fluctuation theorems (FTs) and assisted by new micromanipulation tools, may be another current hot point in scientific researches.

Energy transformation efficiency of a kind of protein molecular motor called kinesin, and so on, were discussed in this paper. They said:

> How efficient is the kinesin motor and how much energy does it dissipate as it moves along the track? The chemical energy released by ATP hydrolysis is about $20k_BT$. (In the world of small systems, the product of Boltzmann constant and temperature is a convenient energy unit.) The motor does about 12 k_BT of work with each step. Thus, the machine's efficiency is roughly 60 % and it dissipates about 650 k_BT per second into its environment.

Note that these estimations were not only concerned with a single molecular motor but also concerned with the statistic average behavior of a large amount of molecular motor. Therefore, the research objects in modern thermodynamics are the collective behavior "of small systems" or the statistic average behavior "of small systems" in a long time interval. The research objects of kinetic mechanism and fluctuation deviation may be really concerned with "small systems". The relationship between thermodynamics and kinetics may be revealed to some extent.

9.7 Conclusion of this Book

> The reversible heat engine must have the highest conversion efficiency of energy from heat to work. That is the Carnot theorem suitable for cyclical processes. In general, the nondissiptaive process or system must have the highest conversion efficiency of energy. That is the extended Carnot theorem suitable for any macroscopic process. The Carnot theorem was macroscopic experiences of human beings summarized by Sadi Carnot in 1824, while the extended Carnot theorem is macroscopic experiences of human beings summarized by the author in the current 21^{st} century. The extended Carnot theorem includes the Carnot theorem, but not vice versa. The

extended Carnot theorem is an extended foundation stone of thermodynamics and the whole view of the thermodynamics discipline is drastically changed.

References

1. Ho S-S, Yan C-S, Liu Z, Mao H-K, Hemley R J. Industrial Diamond Rev [J], 2006, 66: 28-32.
2. Hemley R J, Chen Y-C, Yan C-S. Elements [J], March 2005, 105-108.
3. Yan C-S, Mao H-K, Li W, Qian J, Zhao Y, Hemley R J. Phys Stat Sol (a) [J], 2004, 201: R24-R27.
4. Yan C-S, Vohra Y K, Mao H-K, Hemley R J. PNAS (Proc the National Academy of Sciences) [J], 2002, 99: 1253-12525.
5. Yan C-S, Vohra Y K. Diamond Relat Mater [J], 8: 2022-2031.
6. Nolöng B. Thesis [D],, Faculty of Science, Uppsala University, Sweden, 1983, 691.
7. Wang C-C, Yang G-W. Materials Sci & Eng [J], 2005, R49: 157-202.
8. Hwang N M, Hahn J H, Yoon D Y. J Cryst Growth [J], 1996, 160: 87-97, 98-103 (1996); 1996, 162: 55-68; 1999, 204: 85-90.
9. Piekarczyk W. J Mater Sci [J], 1998, 33: 3443-3453.
10. Piekarczyk W. Cryst Res Technol [J], 1999, 34: 553-563.
11. Lombardi G, Duten X, Hassouni K, et al. J Electrochem Soc [J], 2003, 150(5): C311-C319.
12. Wang J-T. Modern thermodynamics – A whole view of thermodynamics [M]. Shanghai: Fudan Univ Press, 2005.
13. Shen X-F, Hu G, Jiang L. Theory of Dissipative Structures [M]. Shanghai: Shanghai People's Publisher, 1987, 14.
14. Wang J-T. Nonequilibrium Nondissipative Thermodynamics – With Application to Low-Pressure Diamond Synthesis [M]. Heidelberg: Springer, 2002.
15. Wang J-T. Modern Physics Lett B [J], 2002, 16(23 & 24): 885-888.
16. Wang J-T. The Second Law of Thermodynamics in the Current 21st Century (plenary lecture). 2007 National Symp on Thermodynamics and Statistics [C], Yanji, China: 6 Aug 2007.
17. Holman J P. Thermodynamics [M]. 3rd edn. New York: McGraw-Hill, 1980.
18. Pihl T D, Black L K, Schulman B A, et al. J Bacteriology [J], 1992, 174: 137-143.
19. Kniemeyer O, Musat F, Sievert S M, et al. Nature [J], 2007, 449: 898-901.
20. Mitchell P. Science [J], 1979, 206: 1148-1159.
21. Shen R-Q, Gu Q-M. Textbook of Biochemistry [M]. Beijing: High Education Press, 1993.
22. Zubay G. Biochemistry [M]. Massacusetts: Addison-Wesley Publishing Company, 1983, 387-395.
23. Burk D. J Phys Chem [J], 1931, 35: 432.
24. Bustamante C, Liphard J, Ritort F., Physics Today [J], 2005, 48(7): 43-48.

Index