Chapter 1

数与误差

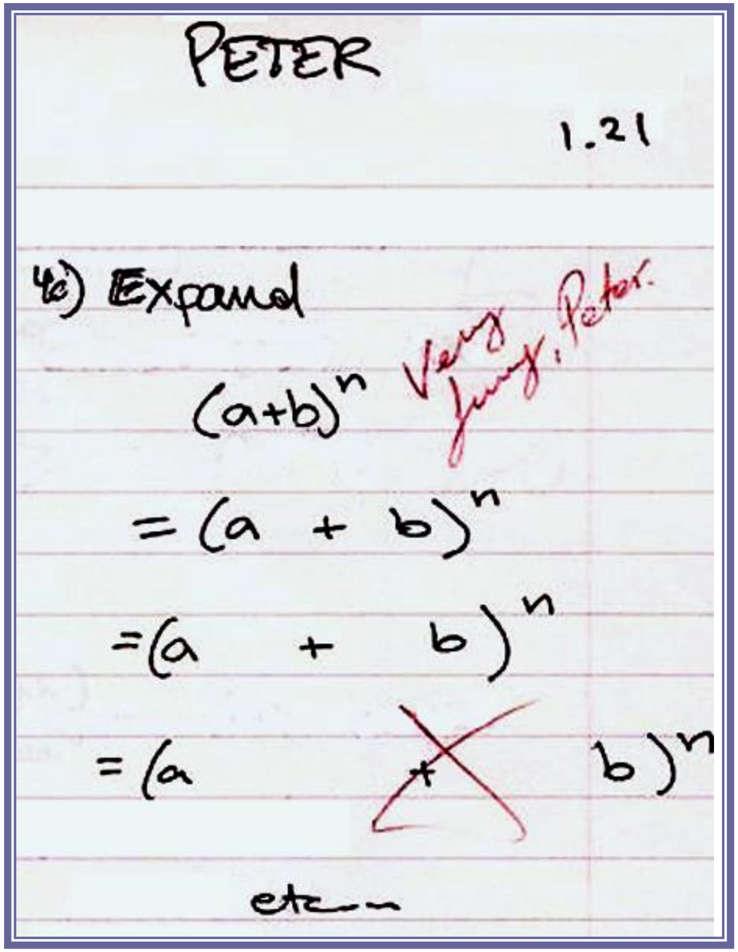
1.1 序：关于“数”的话题

某船上有山羊7头，船上的绵羊数比山羊的2倍还多4只，问船老大今年几岁?

必须做，怎么办？

加、减、乘 、除 ？

这样做数学题？

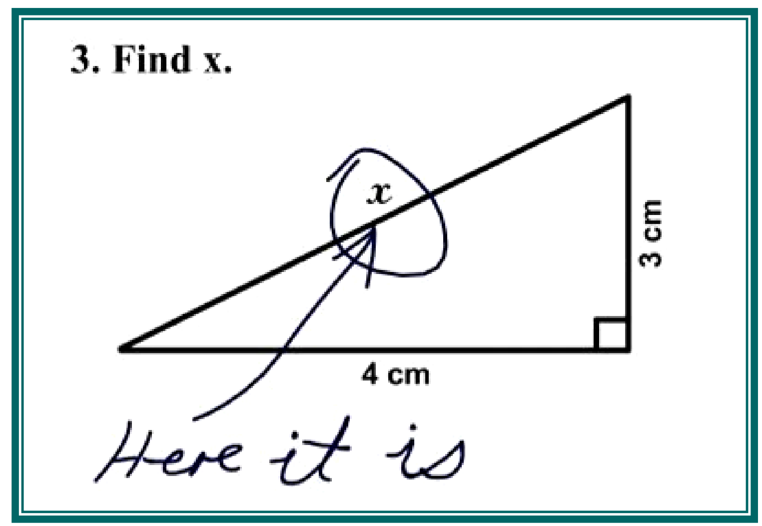


After explaining to a student through various lessons and examples that:

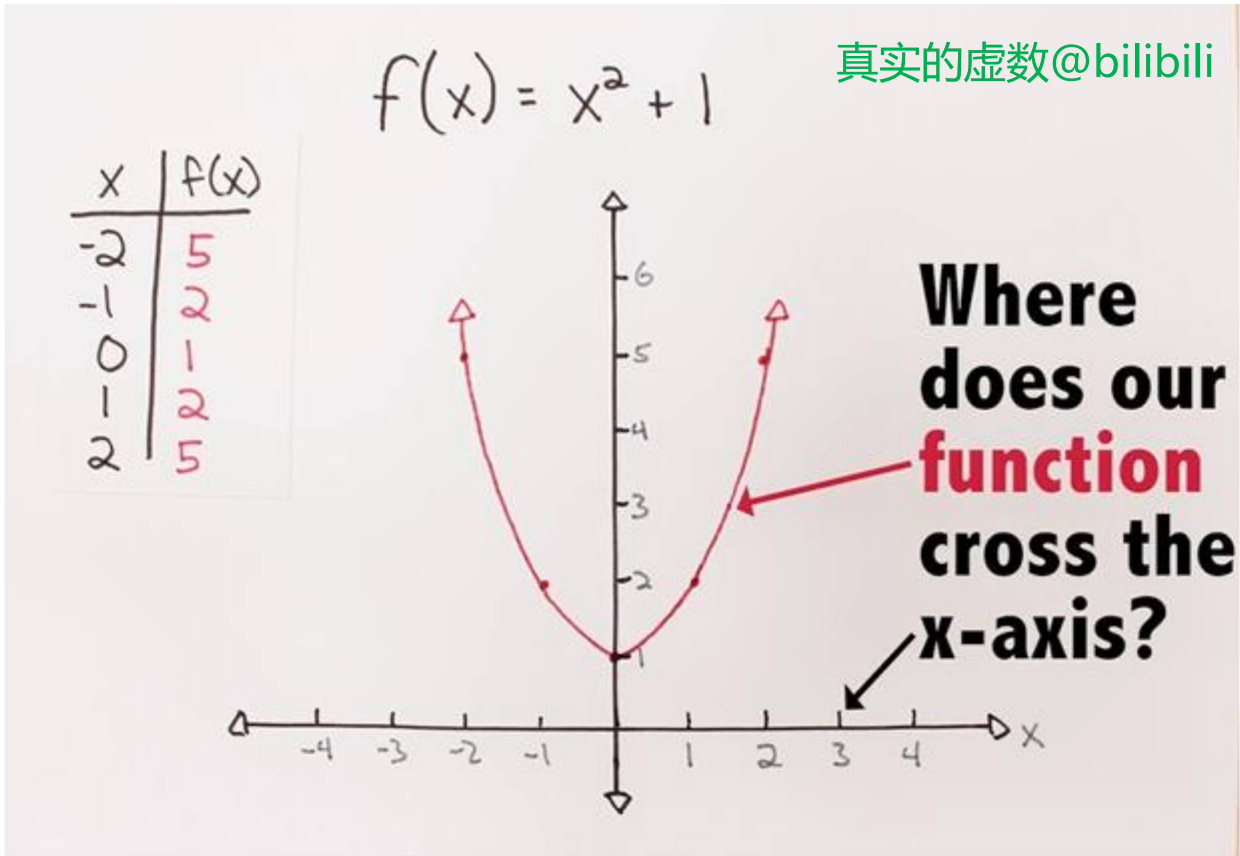
$$\lim_{x \to 8} \frac{1}{x-8} = \infty$$

I tried to check if she really understood that, so I gave her a different example. This was the result:

$$\lim_{x \to 5} \frac{1}{x-5} = 5$$



自然数→整数→小数/分数→实数→复数→？



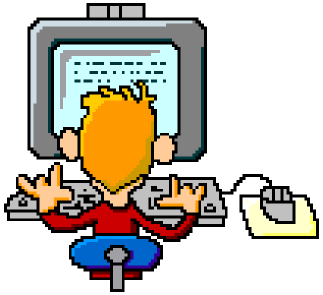
在T_0温度时，1 mol A和1 mol B形成单相溶液。现将温度变化到T_1，测得A在溶液中的活度是0.739，B在溶液中的活度是1。此时系统发生了什么变化?

$\mu = G^0 + RT \ln a$ $\qquad a = \gamma x$

$1 + 1 = 2$ $\qquad 0.739 + 1 = 1.739$

$\ln 0.739 = -0.302\,457\,358 \cdots \cdots$

???



答案：$a_B = 1 \Rightarrow$ 溶液中B饱和、析出

人不能成为“数”的奴隶 !!!

“数”与“量”

数：0, 1, …, 9 以及 ., +, - 等符号的组合

量：具有“单位”的数

〖某次考试试题〗

计算材料成分均匀化过程的时间

“计算”结果：$t = 6.7 \times 10^{-23}$ 秒 ???

“秒”的定义：^{133}Cs在其原子基态的两个超细结构之间辐射周期的 9192631770 倍

一个周期 ~ 1.0878×10^{-10} 秒

时间的最高测量准确度：10^{-15}秒?

1秒=10^3毫秒=10^6微秒=10^9纳秒=10^{12}皮秒=10^{15}飞秒

自然科学工作者需要“量”的概念

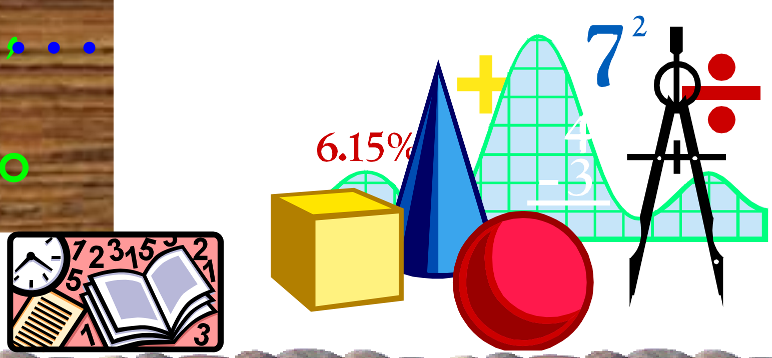
数字游戏不是我们的“专业”

$\pi = 3.14159\ 26535\ 89793\ 23846$
$\quad 26433\ 83279\ 50288\ 41971$
$\quad 69399\ 37510\ 58209\ 74944$
$\quad 59230\ 78164\ 06286\ 20899$
$\quad 86280\ 34825\ 34211\ 70679$
$\quad \ldots\ldots\ldots\ldots\ldots\ldots$

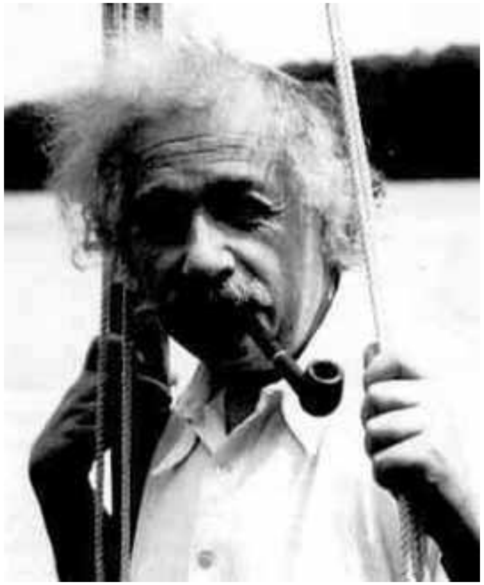
径一而周三

《周髀算经》(约公元前2世纪)

山巅一寺一壶酒，
尔乐苦煞吾，
把酒吃，酒杀尔，
杀不死，乐尔乐。



学会找问题



孔子：　疑是思之始，学之端。

朱熹：　读书无疑者，须教有疑，有疑者却要无疑。

爱因斯坦：　发现问题和系统阐述问题可能比得到解答更为重要。

装模作样是不能发现问题的

测量标尺与海岸线长度

标尺长度 (km)	海岸线长度 (km)
500	2600
100	3800
54	5770
17	8640
0.001	?
$\to 0$	$\to \infty$

分数维数
Fractal Dimension

不是测量误差！
海岸线的维数大于1 ($D \approx 1.26$)

奇妙的数字组合 —— 贾宪三角形

$$(a+b)^n$$

1

1 1

1 2 1

1 3 3 1

1 4 6 4 1

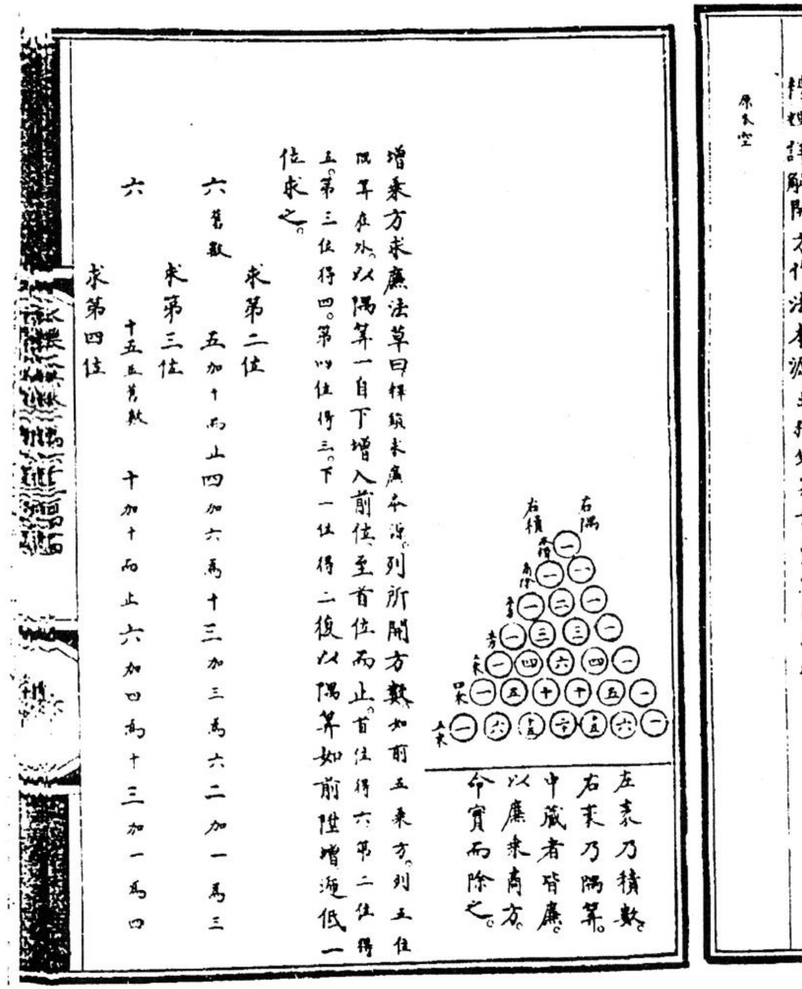
1 5 10 10 5 1

1 6 15 20 15 6 1

1 7 21 35 35 21 7 1

贾宪三角形的相关历史

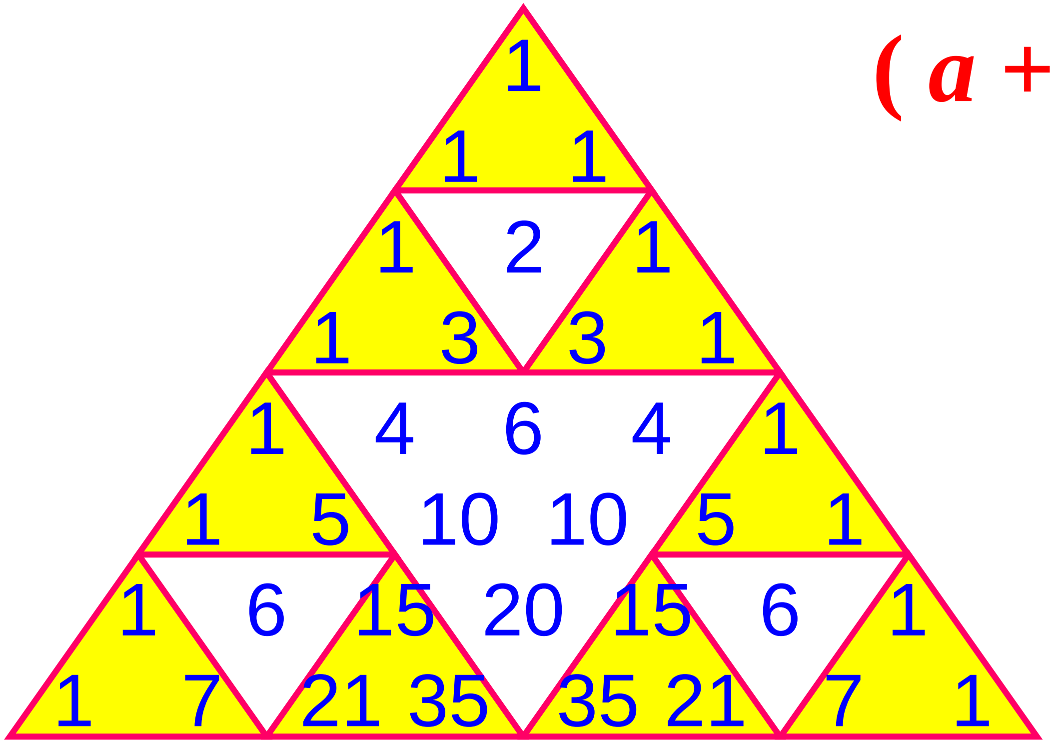
- 11世纪北宋数学家贾宪
  《释锁算术》
  《永乐大典》卷16344
- 13世纪南宋数学家杨辉
  《详解九章算术》
- 14世纪元代数学家朱世杰
  《四元玉鉴》
- 17世纪法国数学家帕斯卡
  （Blaise Pascal, 1623-1662）



《永乐大典》

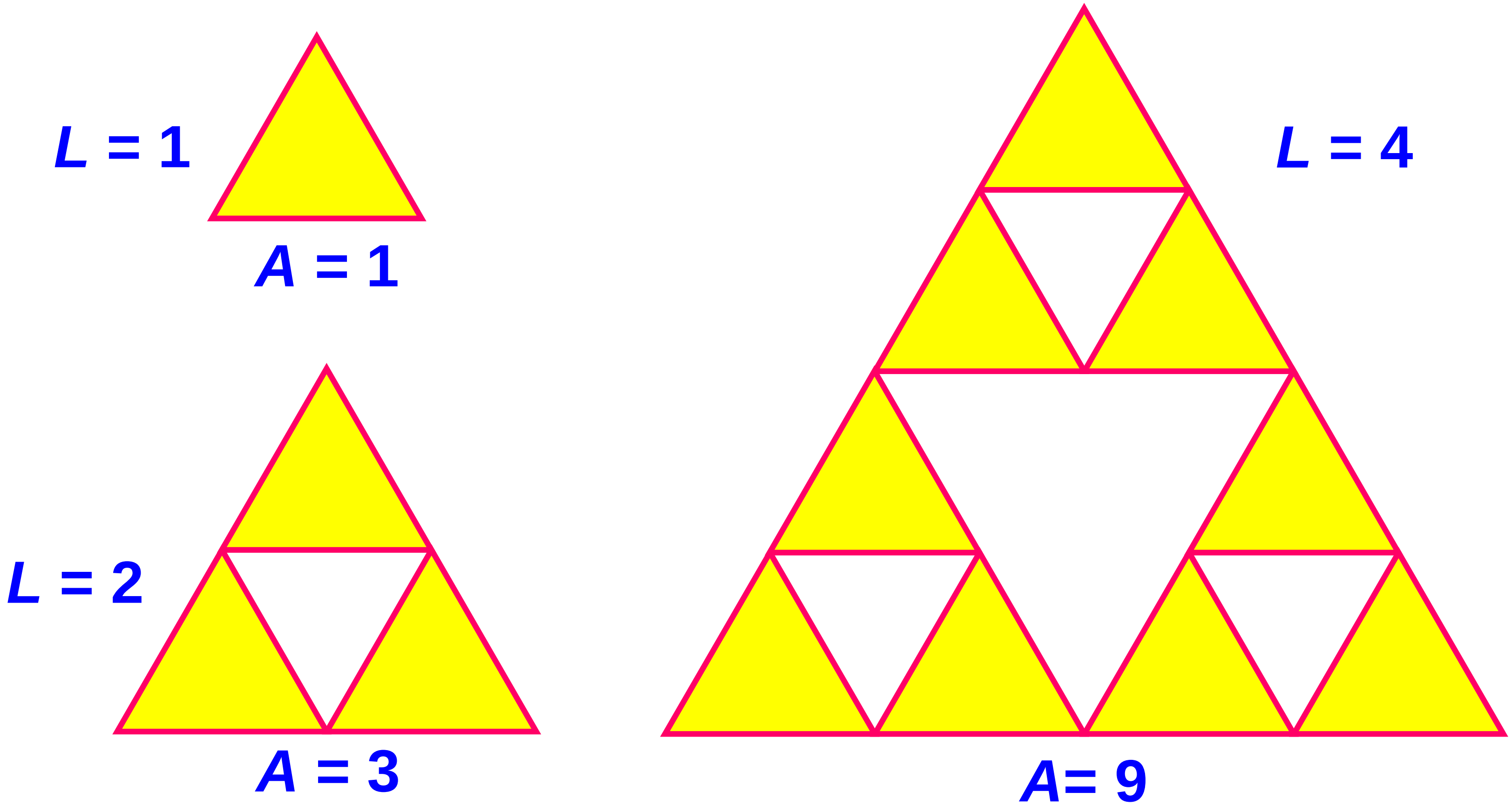
贾宪三角形与分形图案

$(a+b)^n$

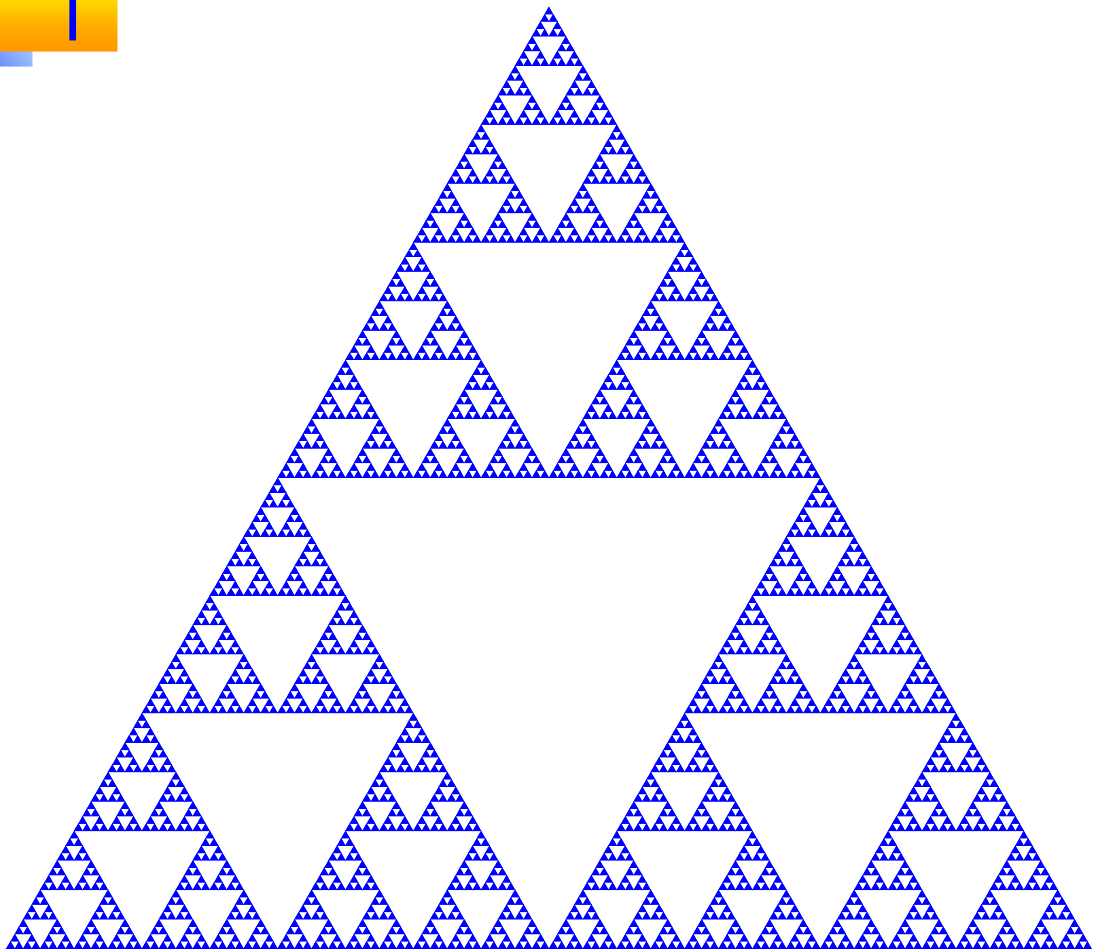


……………………………………

抽象：分形图案



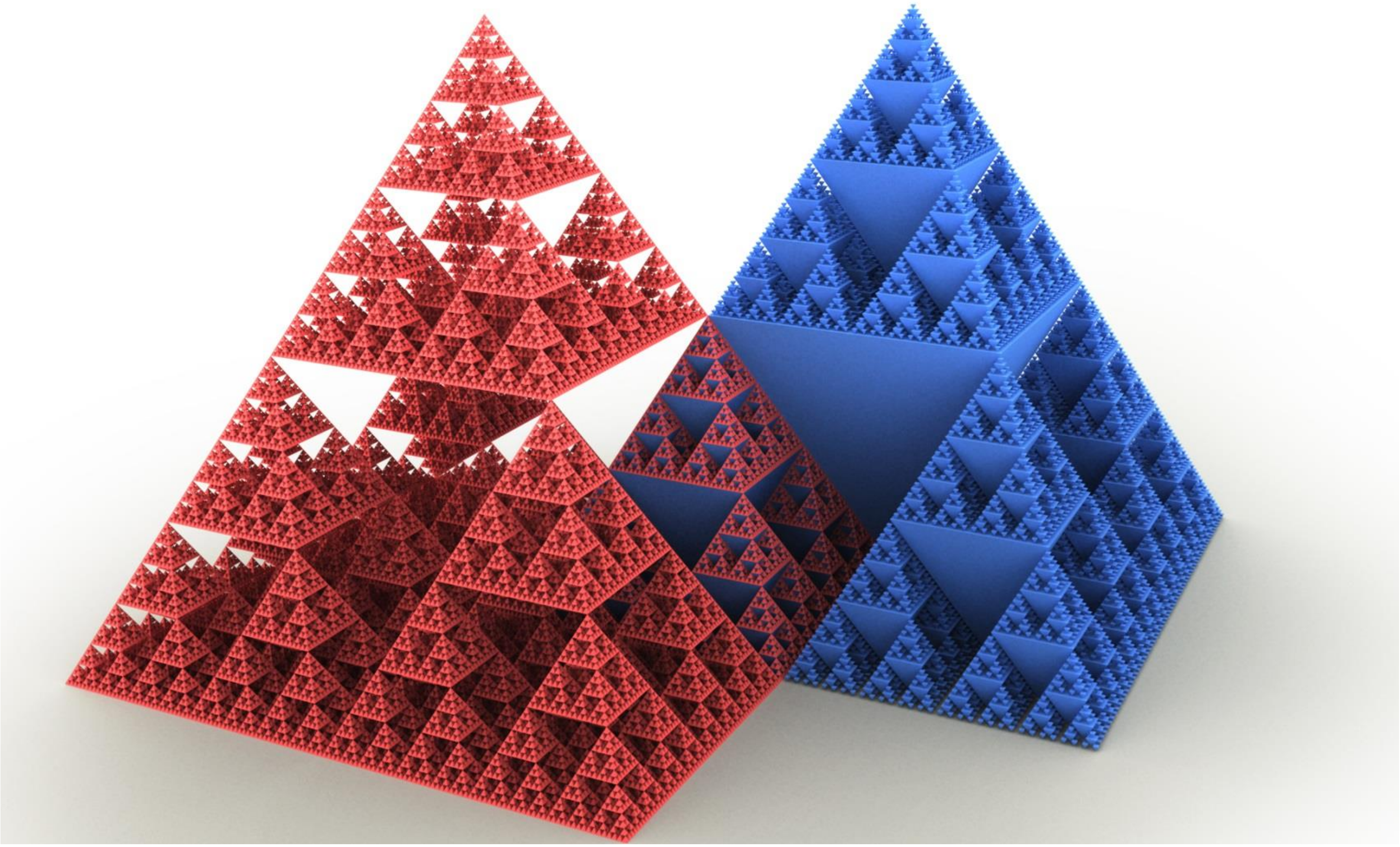
Sierpinski Gaskets（谢尔宾斯基垫圈）



L	*A*
1	1
2	3
4	9
8	27

2^k	3^k

3D 谢尔宾斯基海绵



欧氏平面图形的面积：

$$A = CL^2$$ L：边长，C：比例常数

维数：

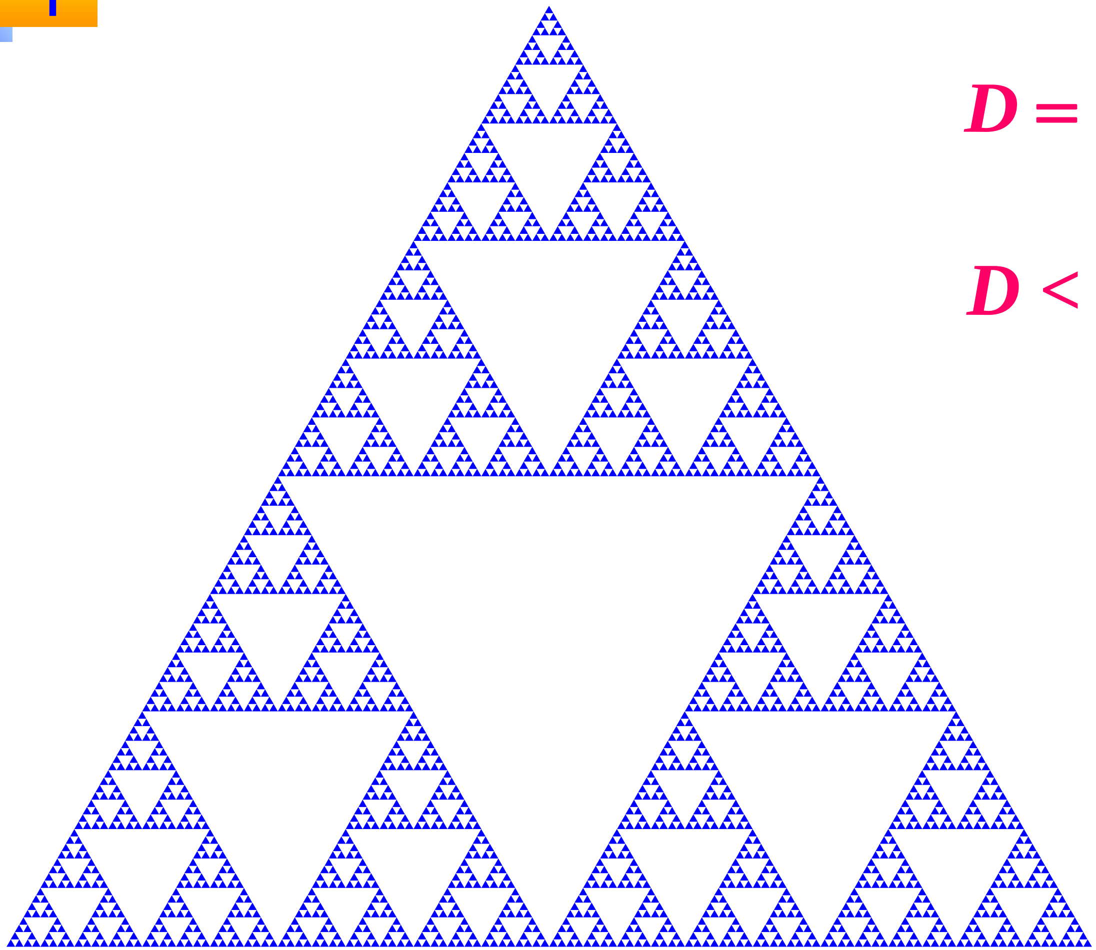
$$D = \frac{d \ln A}{d \ln L} = 2$$ (二维图形)

Sierpinski Gaskets:

$$L = 2^k \qquad A = 3^k$$

$$D = \frac{d \ln A}{d \ln L} = \frac{\ln 3}{\ln 2} \approx 1.585$$ (分数维数)

Sierpinski Gaskets 的 “面积” 测量



$$D = \frac{d\ln A}{d\ln L} = \frac{\ln 3}{\ln 2} \approx 1.585$$

$$D < 2$$

用二维标尺测量，
当标尺 $\rightarrow 0$ 时，
图像面积 $\rightarrow 0$ ！

计算机中的数字

X0	X1 X2......Xn-1 Xn
符号	<--------量位（尾数）--------->

定点数：约定机器中所有数据的小数点位置是固定不变的。由于约定在固定的位置，小数点就不再使用记号“.”来表示。通常将数据表示成**纯小数**或**纯整数**。定点数$X = X0\ X1\ X2 \ldots Xn$ 在定点机中表示如上图($X0$: 符号位，0代表正，1代表负)目前计算机中多采用定点纯整数表示，因此将定点数表示的运算简称为**整数运算**。

浮点数的格式

$$X = M_X * 2^{E_X}$$

浮点数: $X = M_S\ E_S\ E_{m-1} \ldots E_2\ E_1\ M_{-1} M_{-2} \ldots M_{-n}$

IEEE 标准：阶码用移码，基为2；尾数用原码

	符号位	阶码位	尾数数码位	总位数
短浮点数:	1	8	23	32
长浮点数:	1	11	52	64
临时浮点数:	1	15	64	80

浮点数的阶码的位数决定数的表示范围，
尾数的位数决定数的有效精度。

1.2 数值的误差

什么是误差？@维基百科

误差(errors)是实验科学术语。指测量结果偏离真值的程度。对任何一个物理量进行的测量都不可能得出一个绝对准确的数值，即使使用测量技术所能达到的最完善的方法，测出的数值也和真实值存在差异，这种测量值和真实值的差异称为误差。（测量误差）

数值计算分为绝对误差和相对误差。

也可以根据误差来源分为系统误差（又称可定误差、已定误差）、随机误差（又称机会误差、未定误差）和毛误差(又称粗差)。

一、按误差来源分类

- 模型误差 万有引力/相对论

  不可避免：对“实际”的简化、抽象、近似

  模型优化：①接近实际，②易于计算，③便于理解

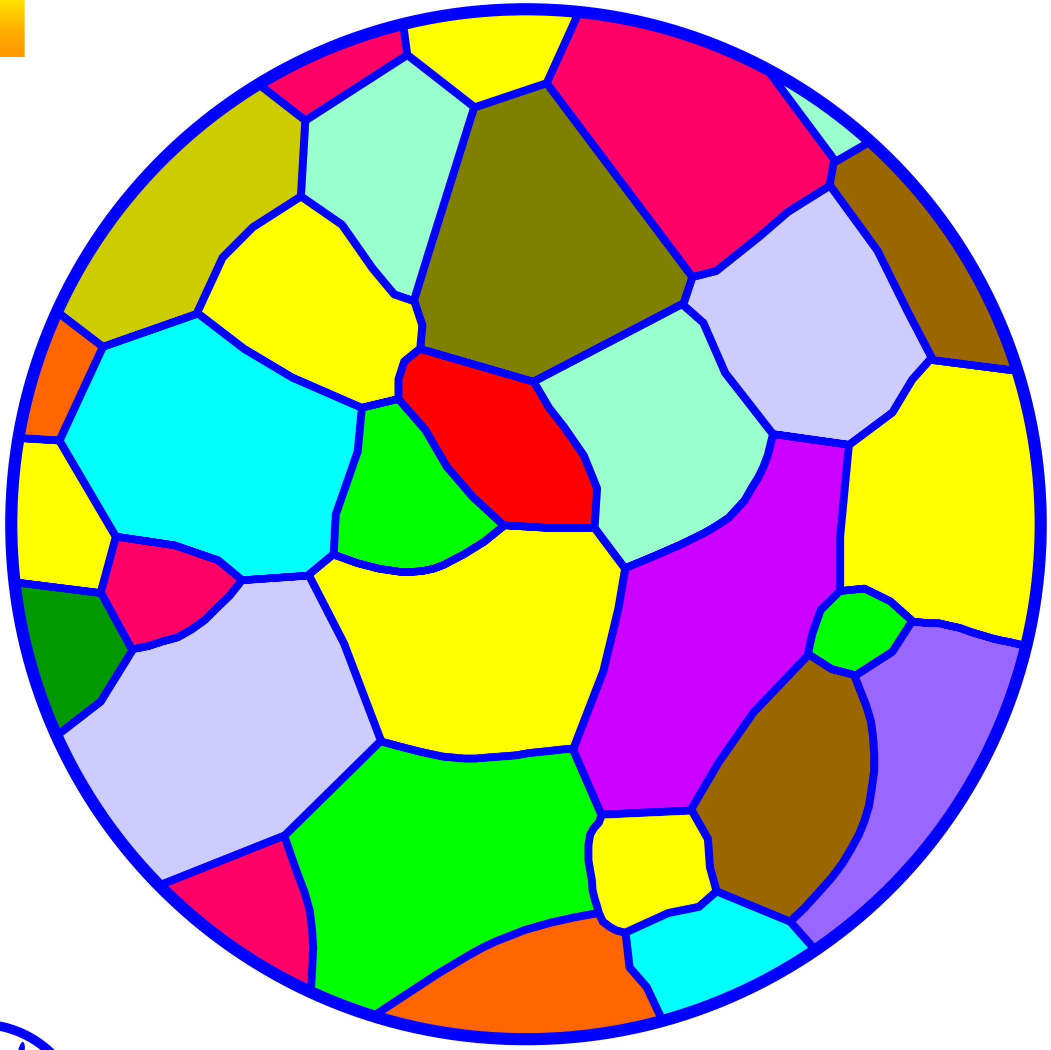
  模型选用不当 ⇒ “致命”错误

- 测量误差 概念在前一页ppt

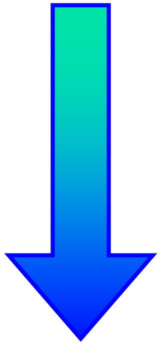
- 计算误差/方法误差

  (参见教材p.2，例1、例2、例3)

[例] 晶粒尺寸测量问题 (模型误差)

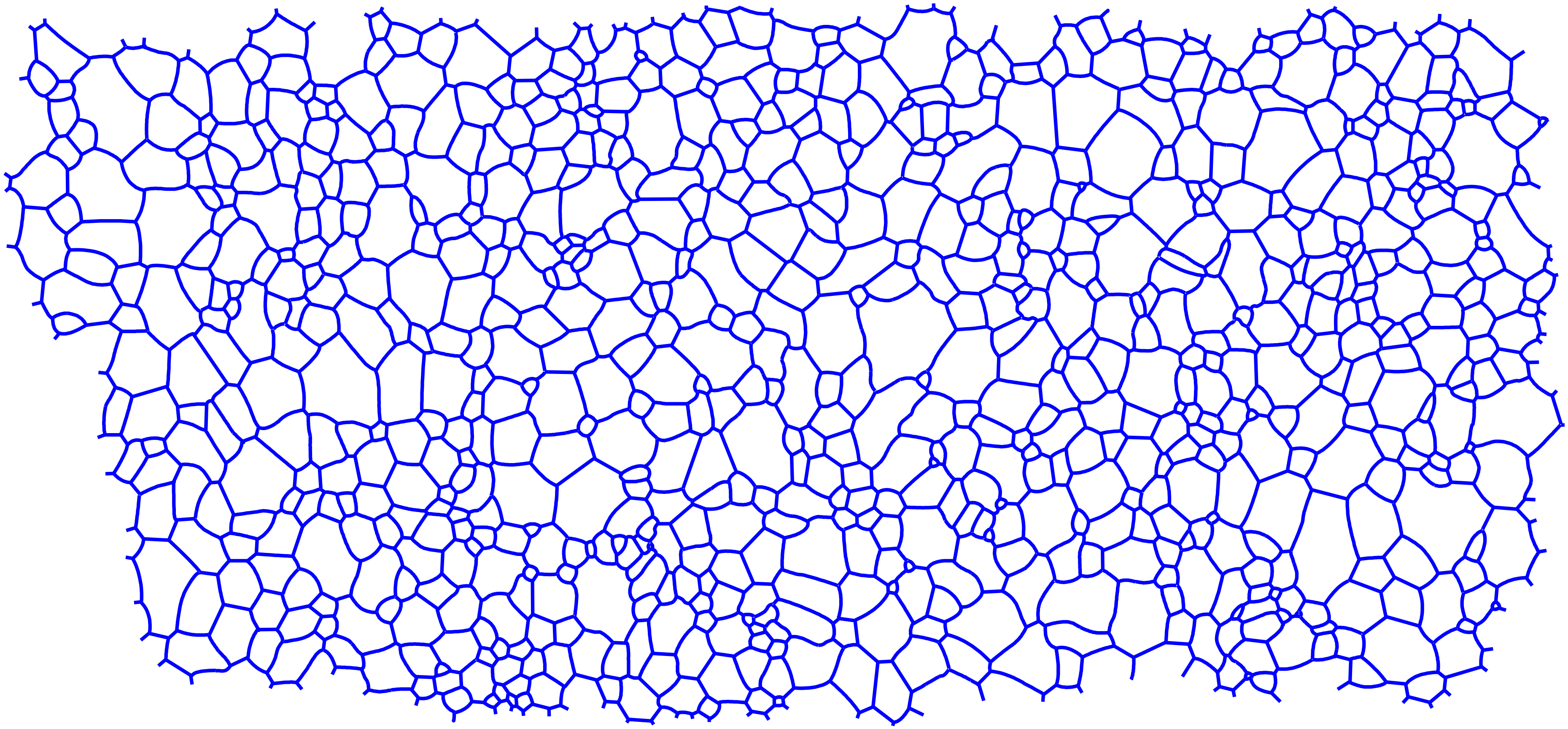


二维图像

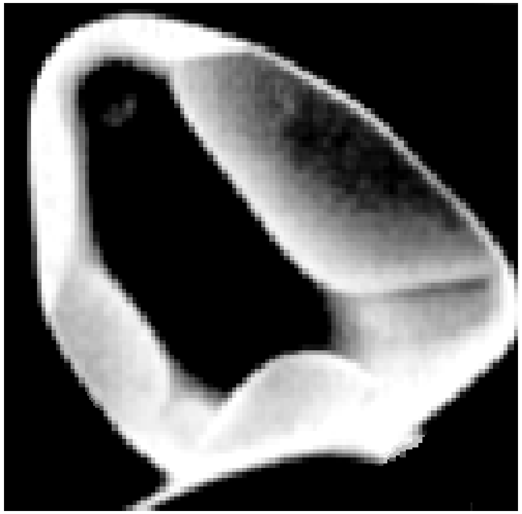


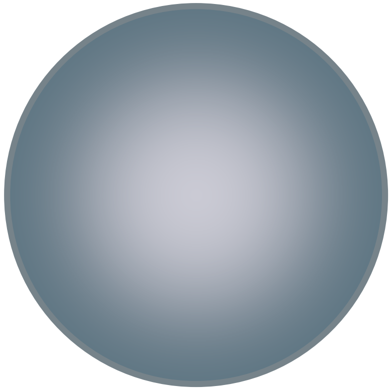
三维晶粒尺寸？

实测二维晶粒图像



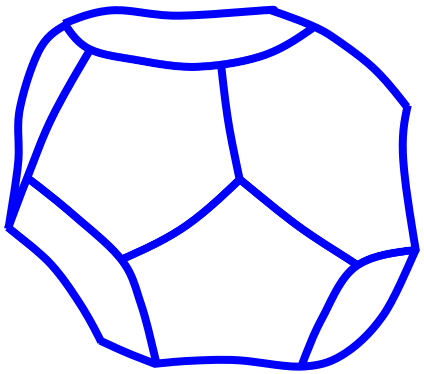
晶粒的几何简化模型



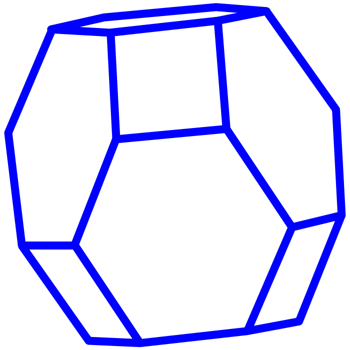


sphere

Spherical model



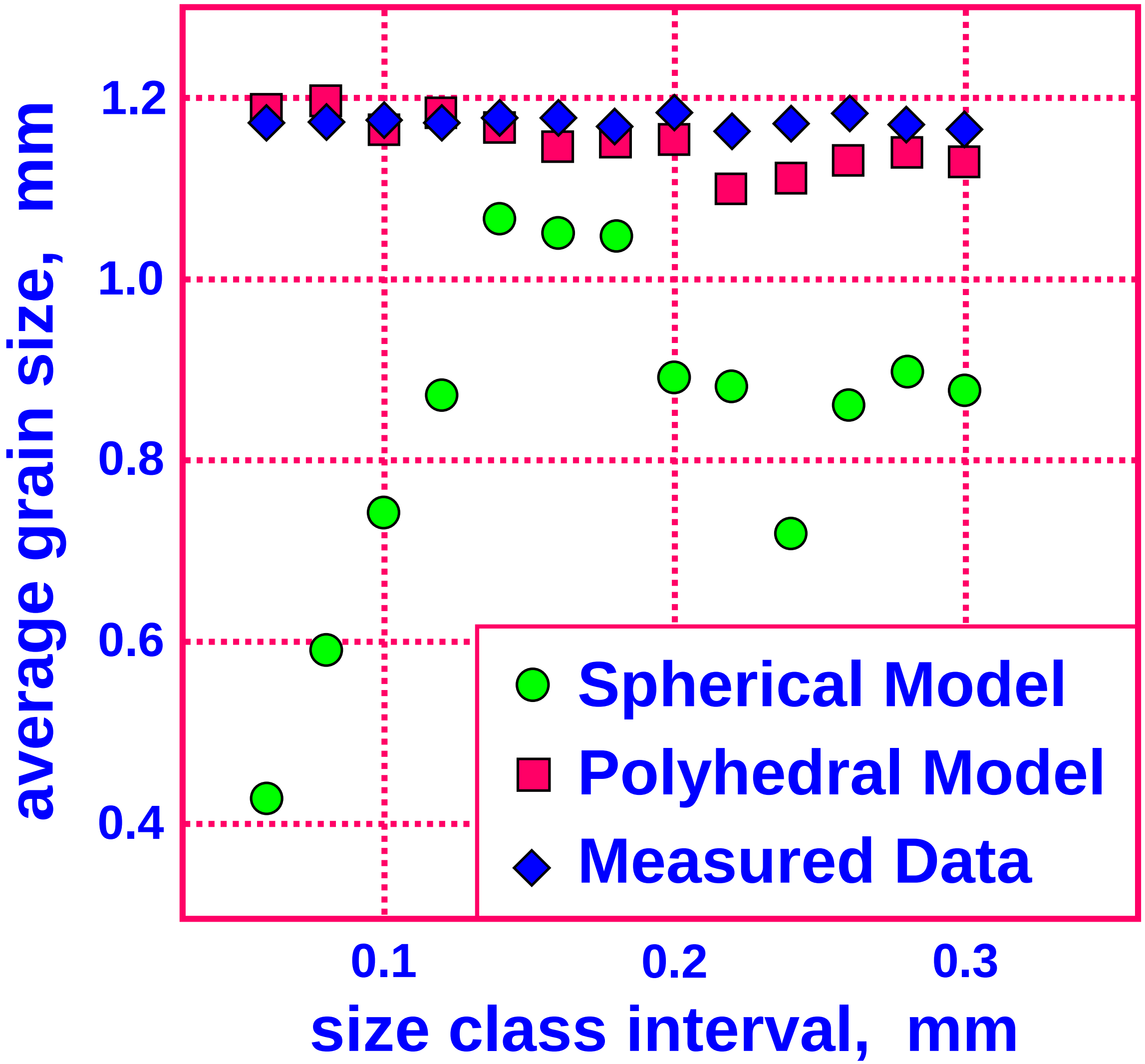
Kelvin's tetradecahedron



tetradecahedron

Polyhedral model

模型误差



1.2
1.0
0.8
0.6
0.4
average grain size, mm
0.1
0.2
0.3
size class interval, mm
Spherical Model
Polyhedral Model
Measured Data

[例] 计算误差

例 1　$y=\arctan 5\,430-\arctan 5\,429$ 的准确值为 0.000 000 033 921 91….但是,若用具有舍入功能的八位计算器,直接按下面过程计算:

$$y=1.570\,612\,2-1.570\,612\,1=0.000\,000\,1,$$

所得近似值很不可靠.

例 2　积分

$$I_n=e^{-1}\int_0^1 x^n e^x \mathrm{d}x \qquad (n=0,1,2,\cdots)$$

的值必定落在区间[0,1]中,而且随着 n 的增大而减小.用分部积分法易得递推关系式

$$I_n=1-nI_{n-1} \qquad (n=1,2,\cdots). \tag{1.1}$$

若在尾数八位的浮点计算机上先算出 $I_0=1-e^{-1}$ 的近似值,然后利用递推式(1.1)依次算出 $I_1,I_2,I_3,\cdots$的近似值,所得结果见表 1-1. I_{12}的近似值超过了 I_0 的近似值,显然是错误的.此后,随着 n 的增大,错误越来越严重.

二、按误差性质(来源)分类

- 随机误差

  可通过统计处理减小

- 系统误差

  模型选用、测量基准，截断误差，舍入误差。

三、按表达方式分类

- 绝对误差：0.2 g，±0.1°C
- 相对误差：0.5 %，0.2 %

参见p.4

+误差限

绝对误差和相对误差

设$x = \pi = 3.1415926\ldots$,

1. 取近似值x^* =3.14, 绝对误差$e^*(x)=x-x^*= 0.0015926\ldots$,

绝对误差限 $|x-x^*|=0.0015926\ldots \leq 0.002=0.2\times 10^{-2}$

2. 取近似值x^* =3.1416, 绝对误差是 0.0000074…,

误差限$|x-x^*|=0.0000074\ldots \leq 0.000008=0.8\times 10^{-5}$

3. 取近似值x^* =3.1415, 绝对误差是 0.0000926…,

$|x-x^*|=0.0000926\ldots \leq 0.0001=0.1\times 10^{-3}$

可见，绝对误差限ε^*不是唯一的。

$e_r(x)=e^*(x)/x$称为x^*的相对误差 → $e_r^*(x)=e^*(x)/x^*$,

$\varepsilon_r^*= |e_r^*(x)|$ 则称为x^*的相对误差限。

有效数字

① 用四舍五入取准确值的前n位x^*作为近似值，则x^*必有n位有效数字。

例如：3.142作为 π的近似值有4位有效数字，而3.141为3位有效数字。

② 有效数字相同的两个近似数，绝对误差不一定相同。

例如：设x_1^*=12345, 设x_2^*=12.345, 两者均有5位有效数字但绝对误差不同。

$$|x-x_1^*|=|x-12345|\le 0.5=1/2\times 10^0$$

$$|x-x_2^*|=|x-12.345|\le 0.0005=1/2\times 10^{-3}$$

③ 把任何数乘以10^p (p = 0, ±1, …)不影响有效位数。

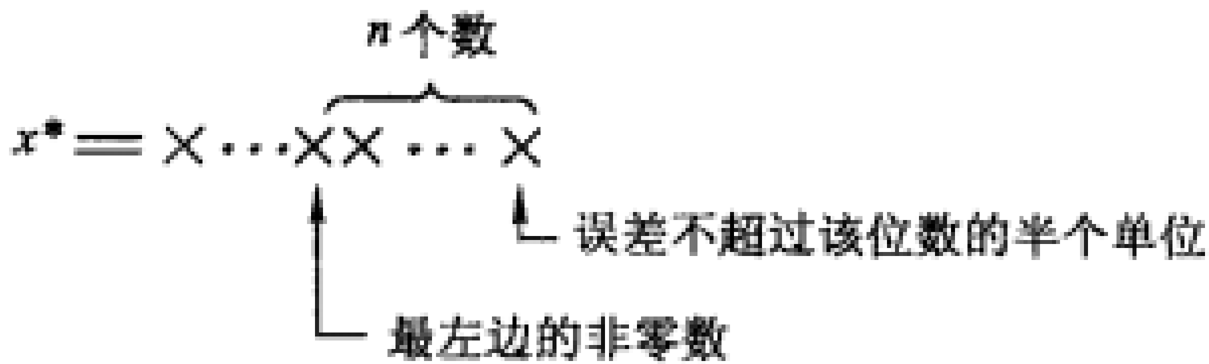


图 1-2

定义 参见 p.6

数值例	有效数字位数	绝对误差限
3.14159	6	0.00001/2
0.01230	4	0.00001/2
2105000	7	1/2
80 kg	2	0.5 kg
80 000 g	5	0.5 g
8 (10^4 g)	1	5 kg

两个近似数 x_1^* 与 x_2^*, 其误差限分别为 $\varepsilon\left(x_1^*\right)$ 及 $\varepsilon\left(x_2^*\right)$, 它们进行加减乘除运算得到的误差限分别为：

$$\varepsilon\left(x_1^* \pm x_2^*\right) \le \varepsilon\left(x_1^*\right) + \varepsilon\left(x_2^*\right)$$

$$\varepsilon\left(x_1^* x_2^*\right) \le \left|x_1^*\right| \varepsilon\left(x_2^*\right) + \left|x_2^*\right| \varepsilon\left(x_1^*\right)$$

$$\varepsilon\left(x_1^* / x_2^*\right) \le \frac{\left|x_1^*\right| \varepsilon\left(x_2^*\right) + \left|x_2^*\right| \varepsilon\left(x_1^*\right)}{\left|x_2^*\right|^2} \left(x_2^* \ne 0\right)$$

例

例 4 测得圆环(图 1-3)外径 $D_1=(10\pm0.05)$ cm,内径 $D_2=(5\pm0.1)$ cm,则其面积 $S=\frac{\pi}{4}(D_1^2-D_2^2)$ 的近似值为

$$S^*=\frac{\pi}{4}[(D_1^*)^2-(D_2^*)^2]$$

$$=\frac{\pi}{4}(10^2-5^2)=\frac{75}{4}\pi\approx58.905\ \text{cm}^2.$$

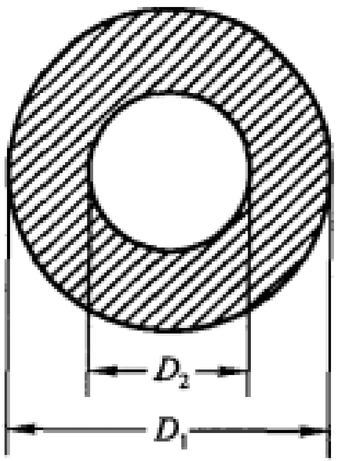


图 1-3

由近似等式(1.8)知,S^* 的绝对误差

$$e^*(S)\approx\frac{\pi}{2}D_1^*e^*(D_1)-\frac{\pi}{2}D_2^*e^*(D_2).$$

现已知 $D_1^*=10$ cm, $D_2^*=5$ cm, $|e^*(D_1)|\leqslant0.05$ cm, $|e^*(D_2)|\leqslant0.1$ cm,故

$$|e^*(S)|\approx\left|\frac{\pi}{2}D_1^*e^*(D_1)-\frac{\pi}{2}D_2^*e^*(D_2)\right|$$

$$\leqslant\frac{\pi}{2}D_1^*|e^*(D_1)|+\frac{\pi}{2}D_2^*|e^*(D_2)|$$

$$\leqslant\frac{\pi}{2}\times10\times0.05+\frac{\pi}{2}\times5\times0.1$$

$$=0.5\pi\approx1.5708\ \text{cm}^2.$$

于是 S^* 的相对误差 $e_r^*(S)$ 满足

$$|e_r^*(S)|=\left|\frac{e^*(S)}{S^*}\right|\leqslant\frac{1.5708}{58.905}<0.027=2.7\%.$$

故若取 $S^*=58.905$ cm^2 作为圆环面积的近似值,则其绝对误差不超过 1.5708 cm^2,相对误差小于 2.7%,此时 S^* 至少具有 1 位有效数字.

数值运算中的误差估计

函数： $y = f(x)$

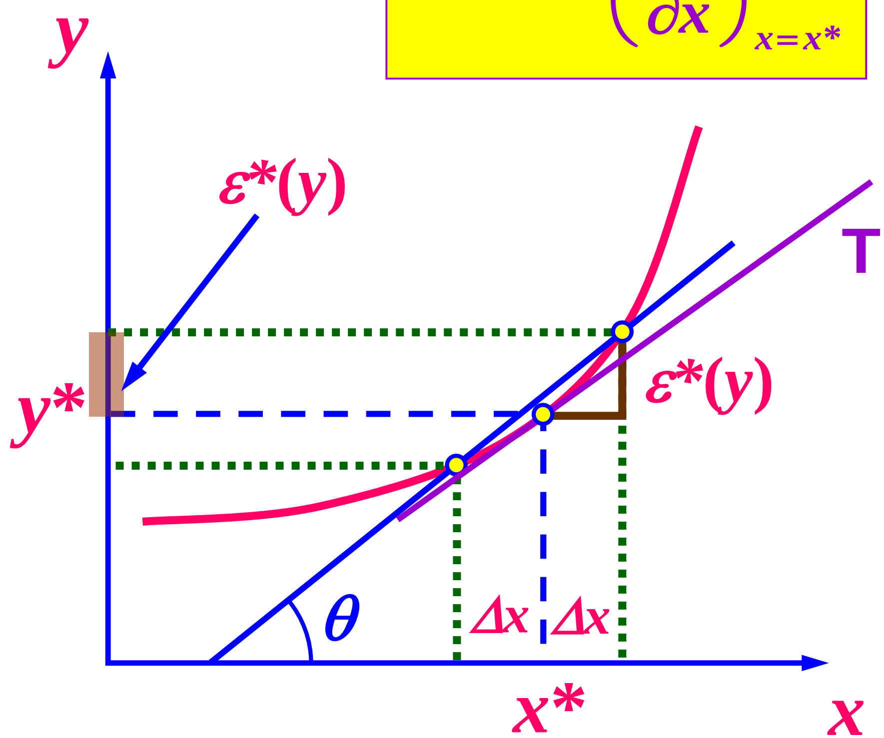
测得 x^*，计算得到 y^*

若 x^* 的测量误差 = $\pm\Delta x$

则 y^* 的绝对误差

$$\varepsilon^*(y) = \Delta x \tan\theta$$

$$\tan\theta \approx \left(\frac{\partial f}{\partial x}\right)_{x=x^*}$$



所以:

$$\varepsilon^*(y) = \left(\frac{\partial f}{\partial x}\right)_{x=x^*} \Delta x$$

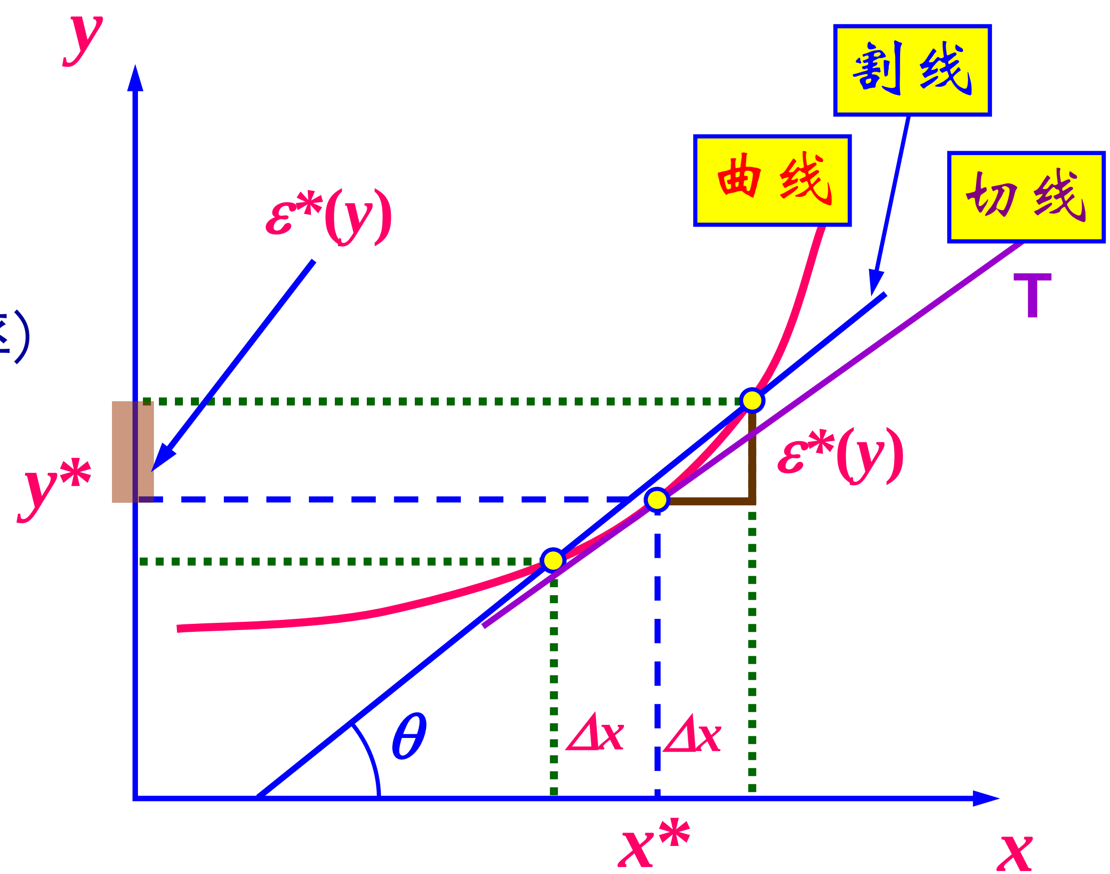
误差估计: “方法” 和 “思想”

方法

用切线代替割线
(切线 ← 曲线斜率)

思想

$\Delta x \to 0$, 切线=割线
导数的定义
误差是 “小量”



思考

什么时候需用割线代替切线？

一、稳定性好的算法

求解 $x=\left(\frac{\sqrt{2}-1}{\sqrt{2}+1}\right)^3$

序号	算式	计算结果	
		$\sqrt{2}\approx 7/5$	$\sqrt{2}\approx 17/12$
1	$(\sqrt{2}-1)^6$	$\left(\frac{2}{5}\right)^6=0.004096$	$\left(\frac{5}{12}\right)^6=0.005233$
2	$99-70\sqrt{2}$	1	$-\frac{1}{6}=-0.166667$
3	$\left(\frac{1}{\sqrt{2}+1}\right)^6$	$\left(\frac{5}{12}\right)^6=0.005233$	$\left(\frac{12}{29}\right)^6=0.005020$
4	$\frac{1}{99+70\sqrt{2}}$	$\frac{1}{197}=0.005076$	$\frac{12}{2378}=0.005046$

近似值之和的绝对误差等于各近似值绝对误差的代数和。

因此在实际计算中，二、应尽量设法避开相近数的相减。

如：例1的改进 @p.9-10

$x \to 0$, $\cos x \to 1$； 求$1-\cos x$的值？

可以用 $1-\cos x = 2\sin^2\left(\dfrac{x}{2}\right)$

近似值之积的相对误差等于相乘各因子的相对误差的代数和。

三、应避免让绝对值太小的数作为除数。

四、防止大数“淹没”小数的现象发生。

1.4 实验数据的描述与处理

实验数据，如：(x_1, y_1)，(x_2, y_2)，……，(x_n, y_n)

1、需要用图表描述

- 图、表的选用 (用图，还是用表)
- 图、表的形式 (图的坐标轴，表的项目)

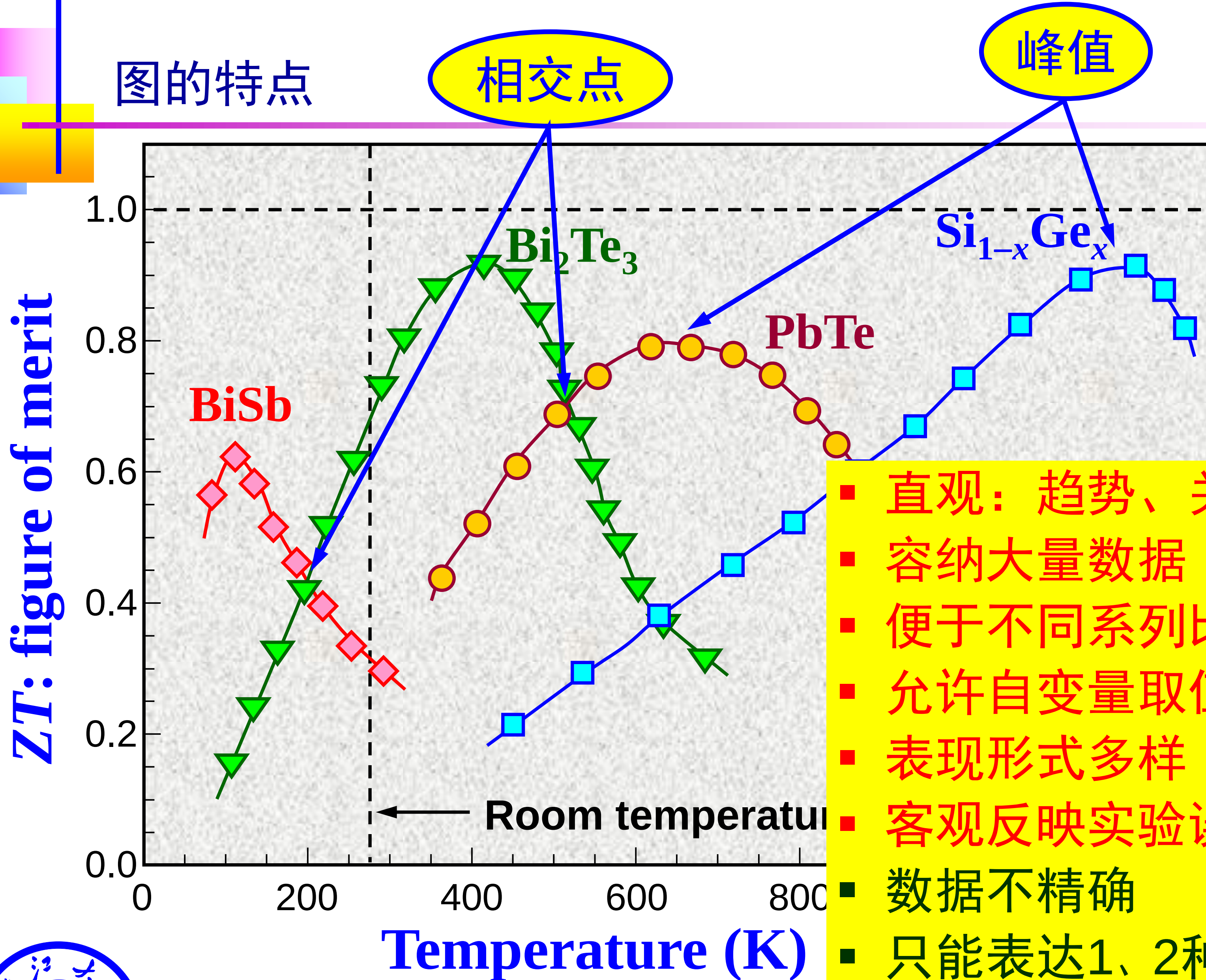
2、离散性数据

- 图中数据点之间的连线问题
- 离散数据的曲线拟合问题

3、包含各种误差

- 误差的分析
- 误差的消除 (减小)

图的特点



- 直观：趋势、关键值
- 容纳大量数据
- 便于不同系列比较
- 允许自变量取值不同
- 表现形式多样
- 客观反映实验误差
- 数据不精确
- 只能表达1、2种性能

表的特点

element / property	Iron	Samarium
electronic structure	$3d^6\ 4s^2$	$4f^6\ 6s^2$
valence	2, 3	2, 3
crystal structure	bcc	bcc
density (g/cm^3)	7.86	7.53
disilicide structure	P4/mmm	Imma

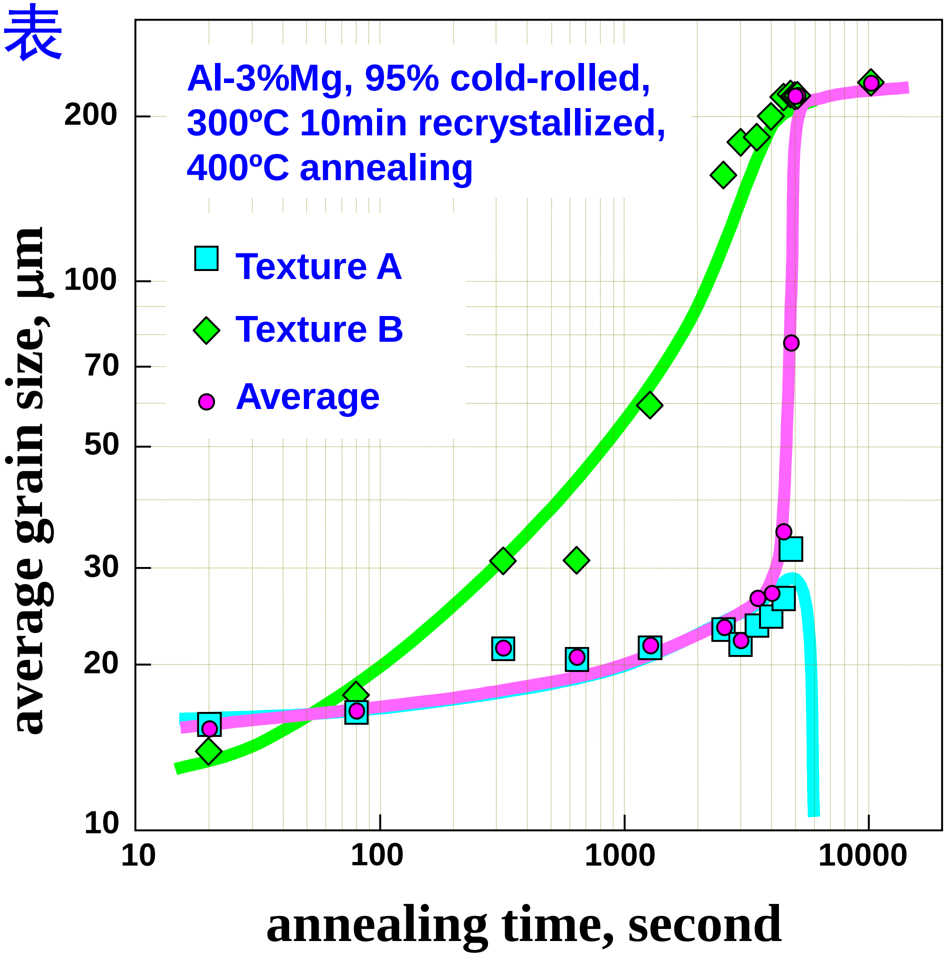
- ❑ 数据“确切”
- ❑ 允许“非数值”数据
- ❑ 性质、量纲多样性

- ❖ 不直观 (对数据表格而言)
- ❖ 数据容量小
- ❖ 表现形式单一

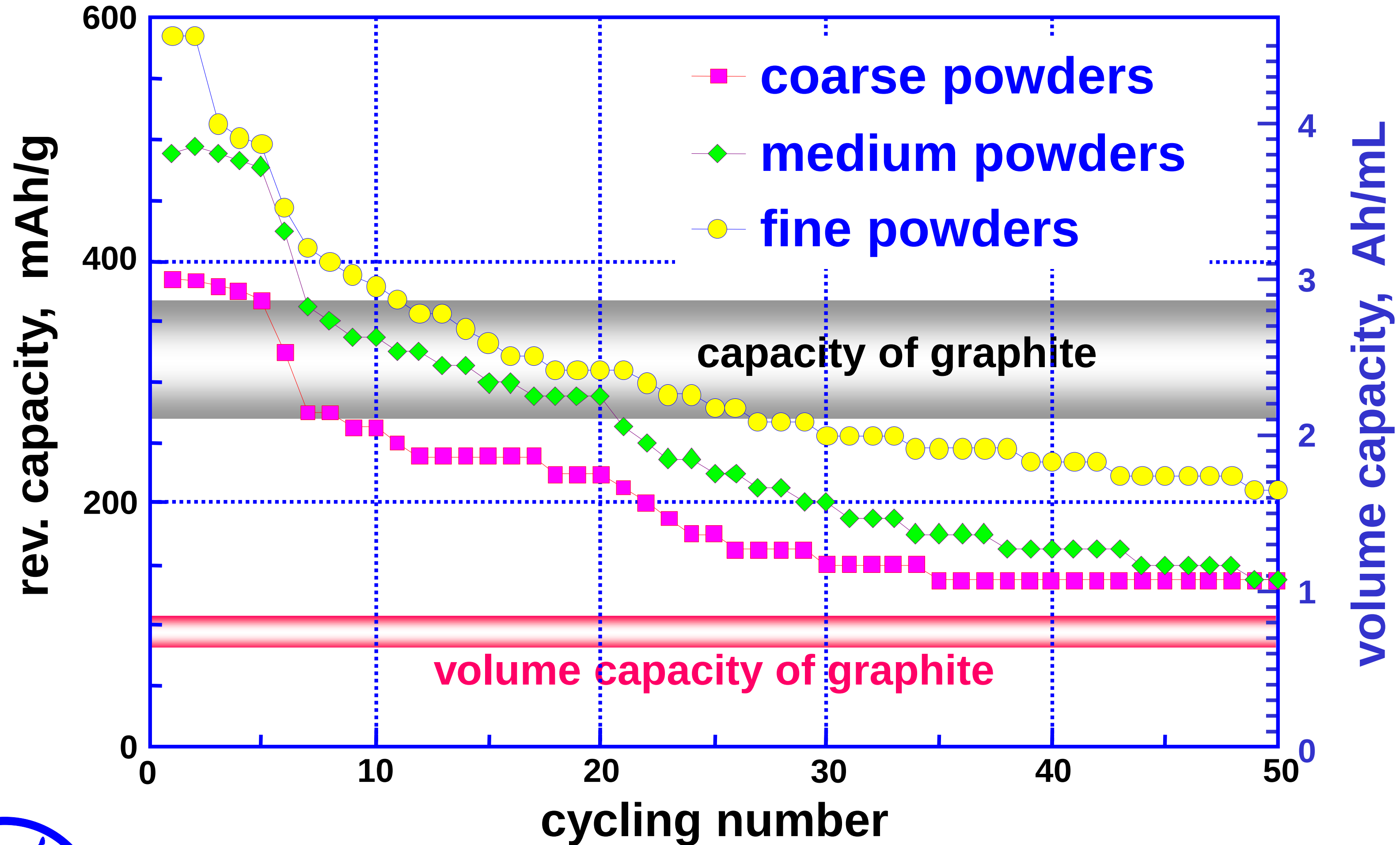
图或表的选择一般原则

- 涉及非数值数据，只能用表
- 涉及大量数据，只能用图
- 为直观确定峰值，用图
- 为直观分析趋势，用图
- 为相互比较，用图
- 为拟合分析，用图

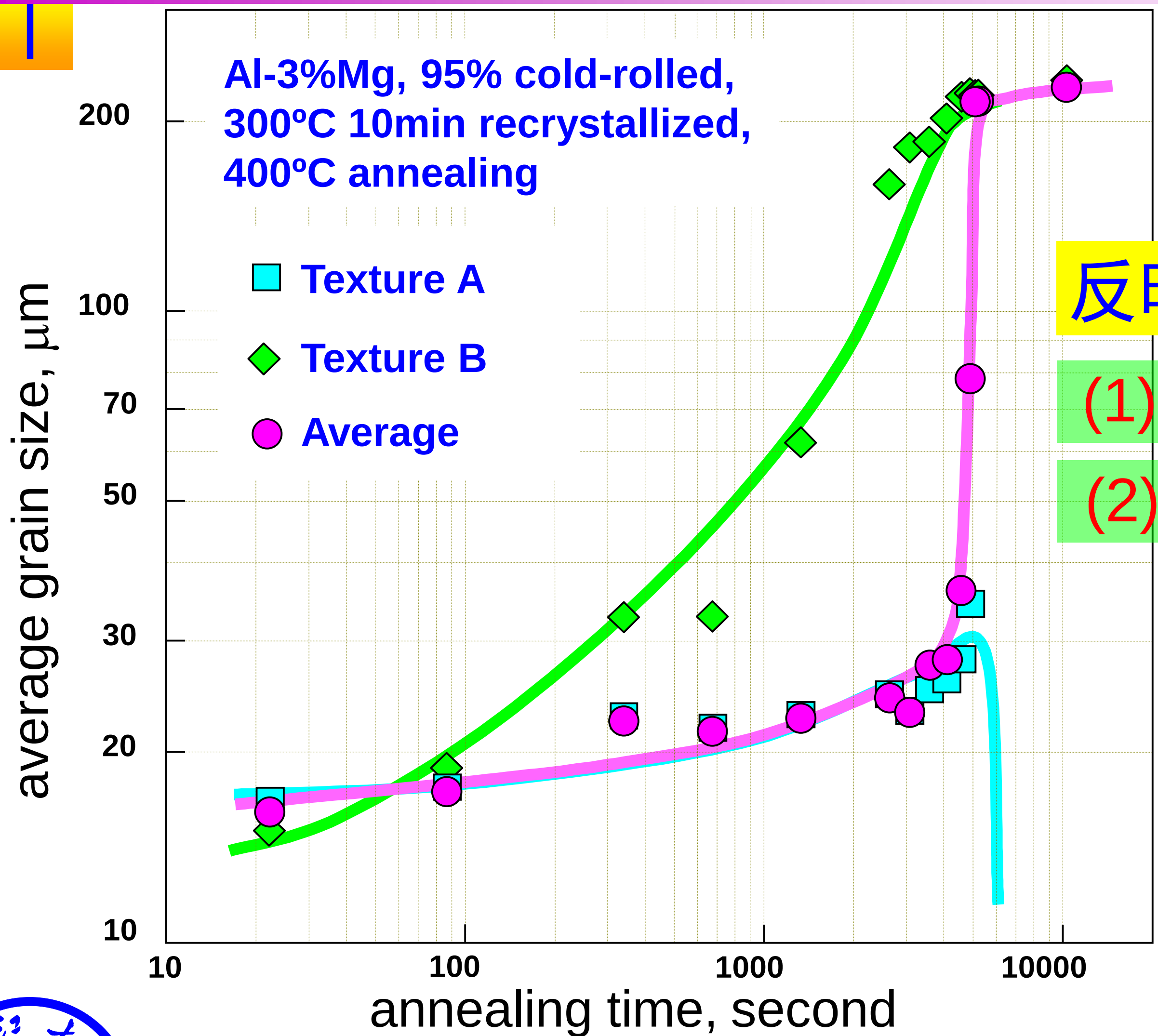
- 实验条件，用表
- 样品编号，用表
- “精确”数据，用表



曲线图的坐标轴 —— 线性坐标轴



曲线图的坐标轴 —— 对数坐标轴



反映物理规律

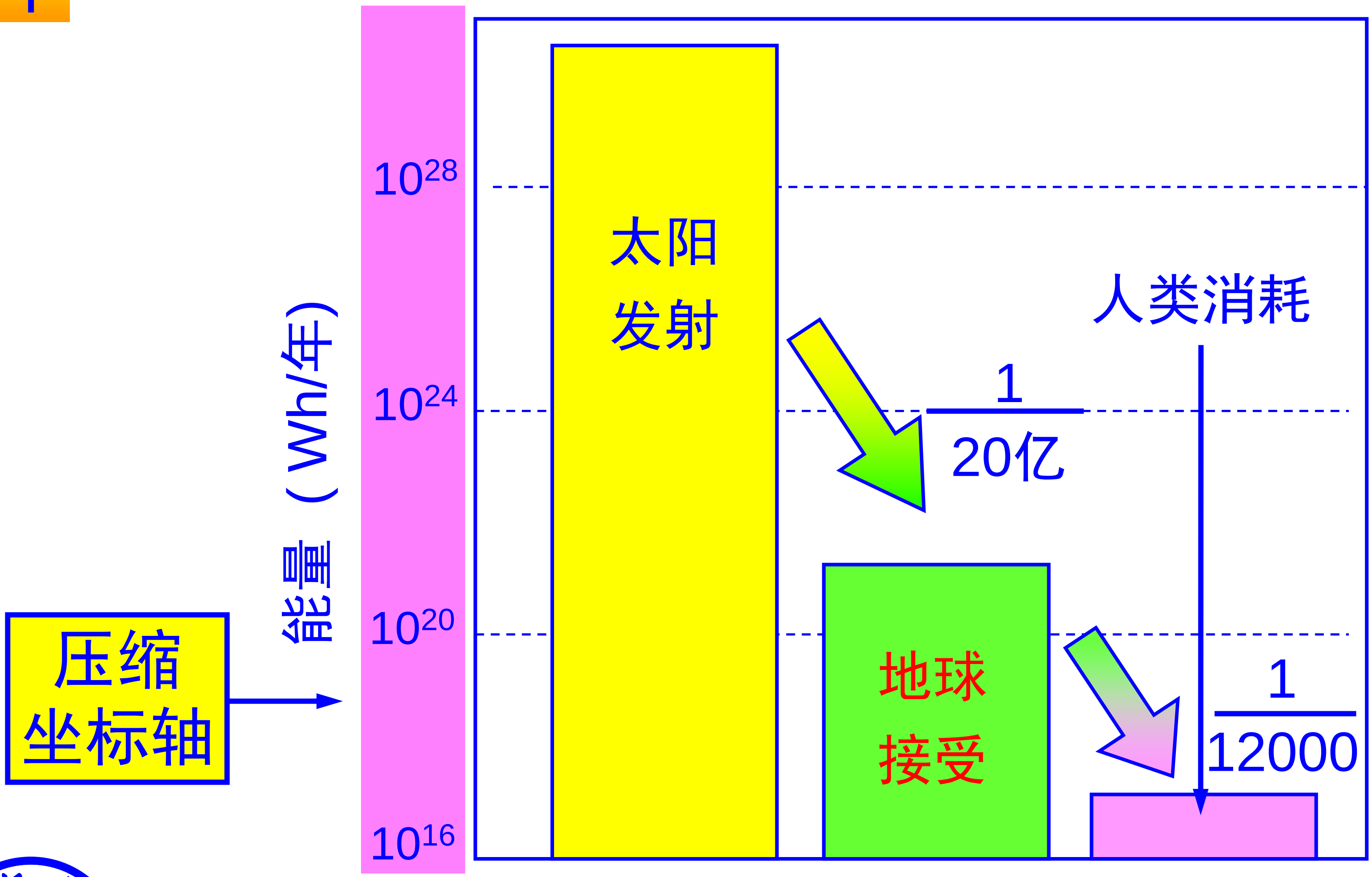
(1) 动力学过程, **ln(*t*)**

(2) 对数相关关系

$$R \sim t^{\eta}$$

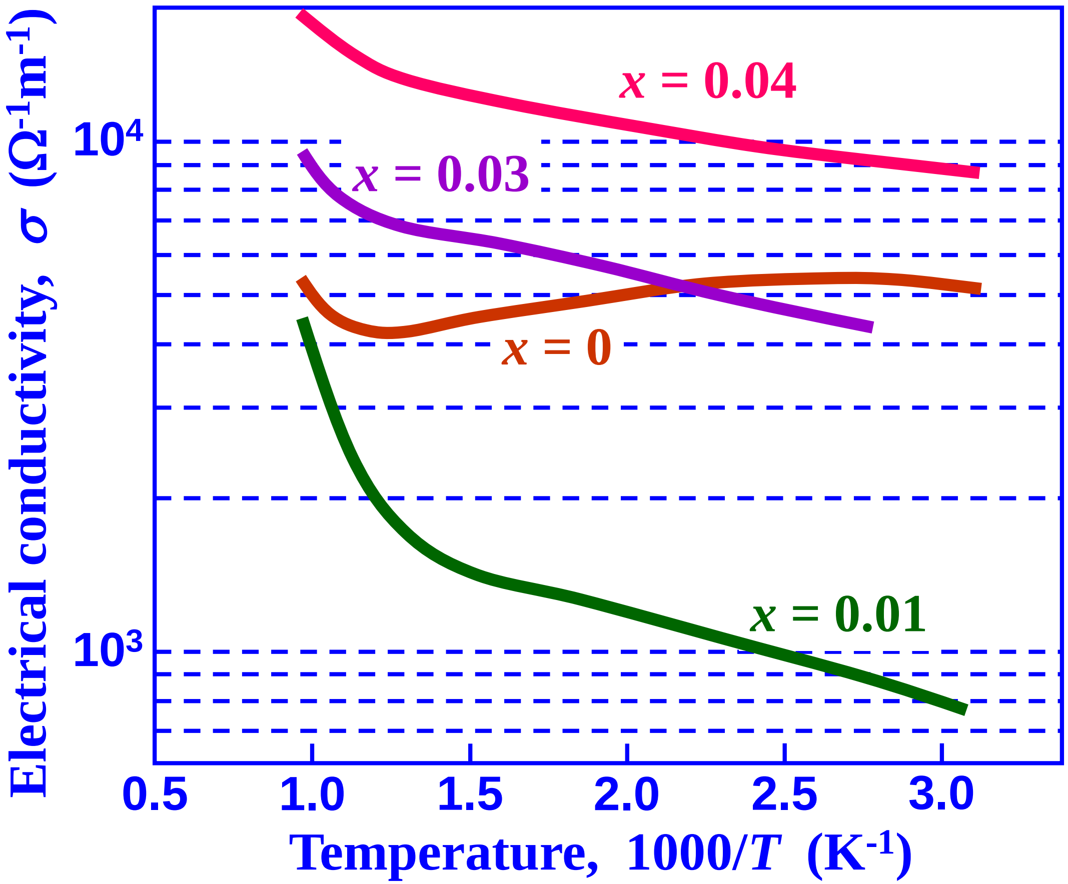
$$\ln R \sim \eta \ln t$$

曲线图的坐标轴 —— 对数坐标轴



曲线图的坐标轴 —— 其他坐标轴

倒数坐标轴、指数坐标轴：反映物理规律



β-$Fe_{1-x}Co_xAl_{0.06}Si_{1.94}Cu_{0.01}$

半导体电导

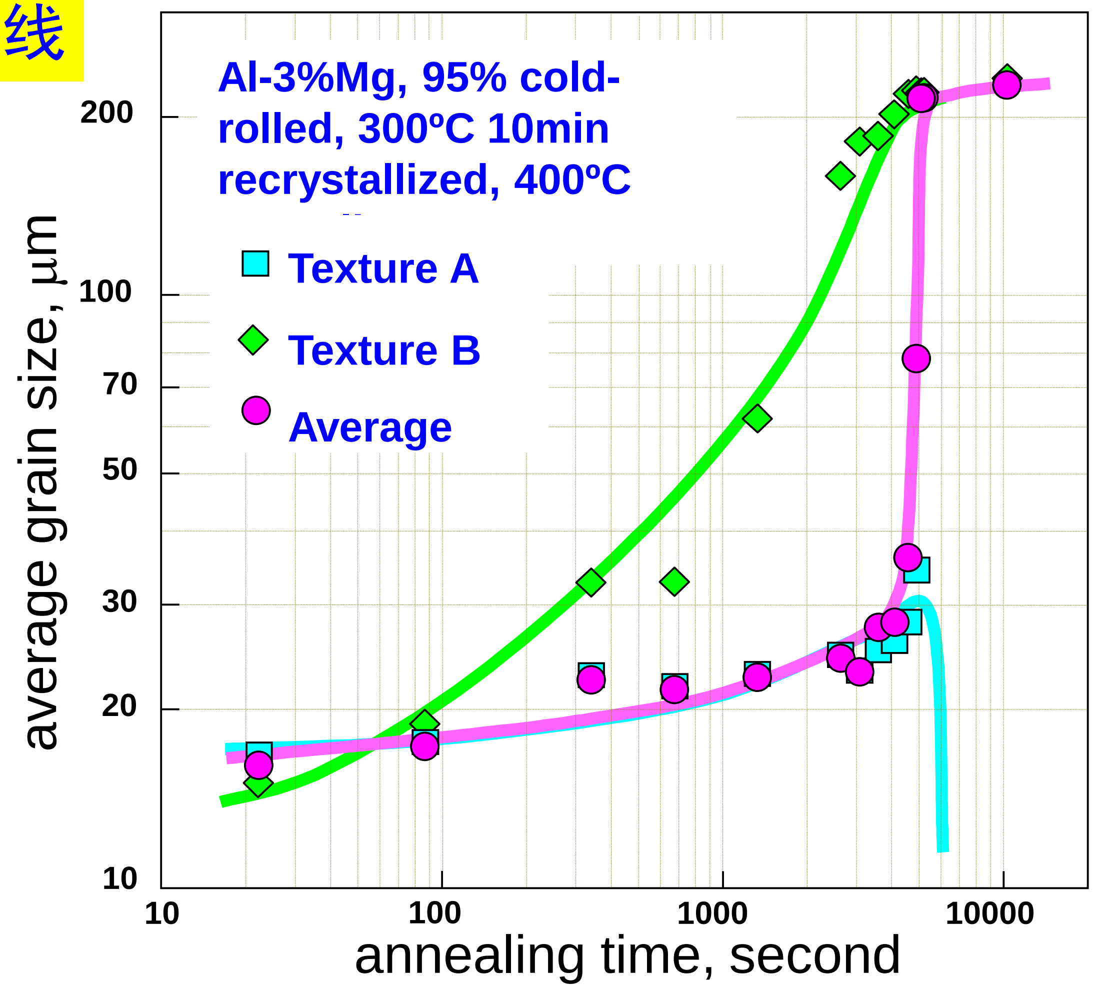
$\sigma \sim \exp(-E/kT)$

Arrhenius定律

$D = D_0\exp(-Q/RT)$

离散数据点之间的连线

首选物理模型拟合曲线

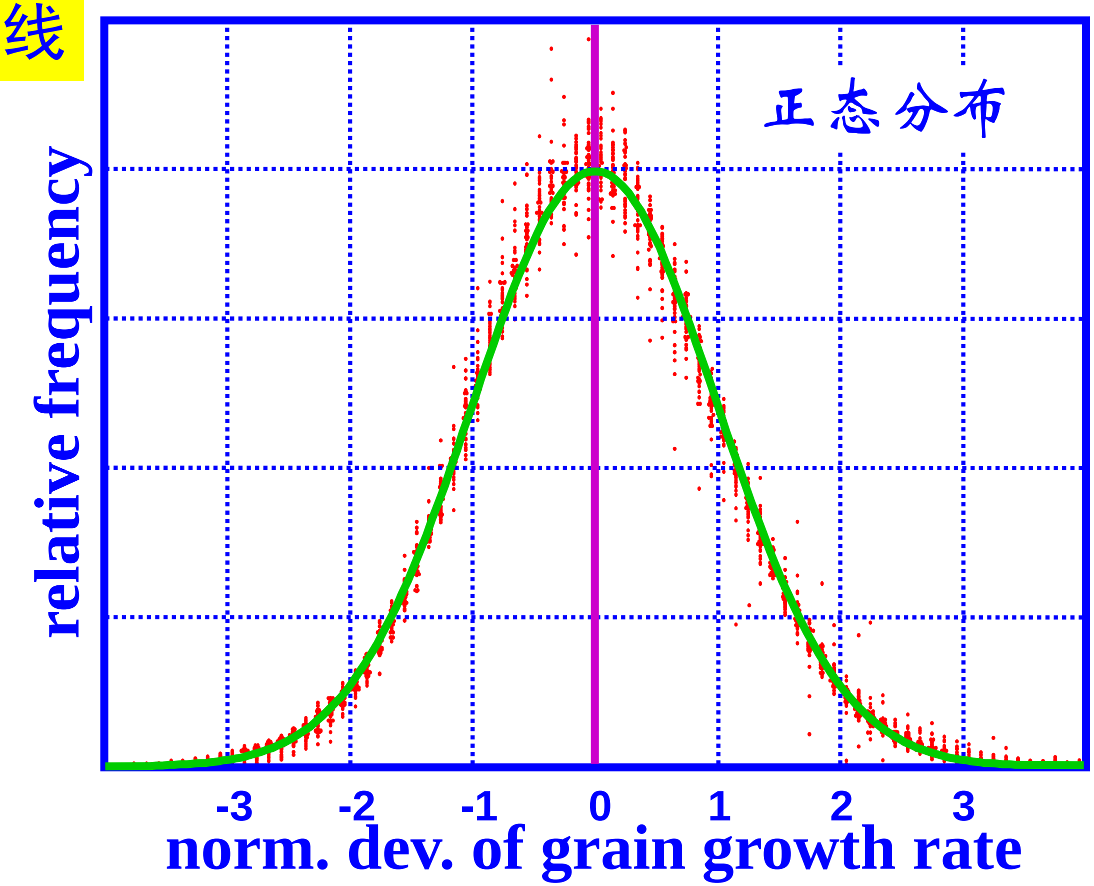


离散数据点之间的连线

首选物理模型拟合曲线

次选常用数理函数曲线

$$y=\frac{1}{\sqrt{2\pi\sigma}}\exp\left\{-\frac{x^2}{2\sigma^2}\right\}$$



离散数据点之间的连线

首选物理模型拟合曲线

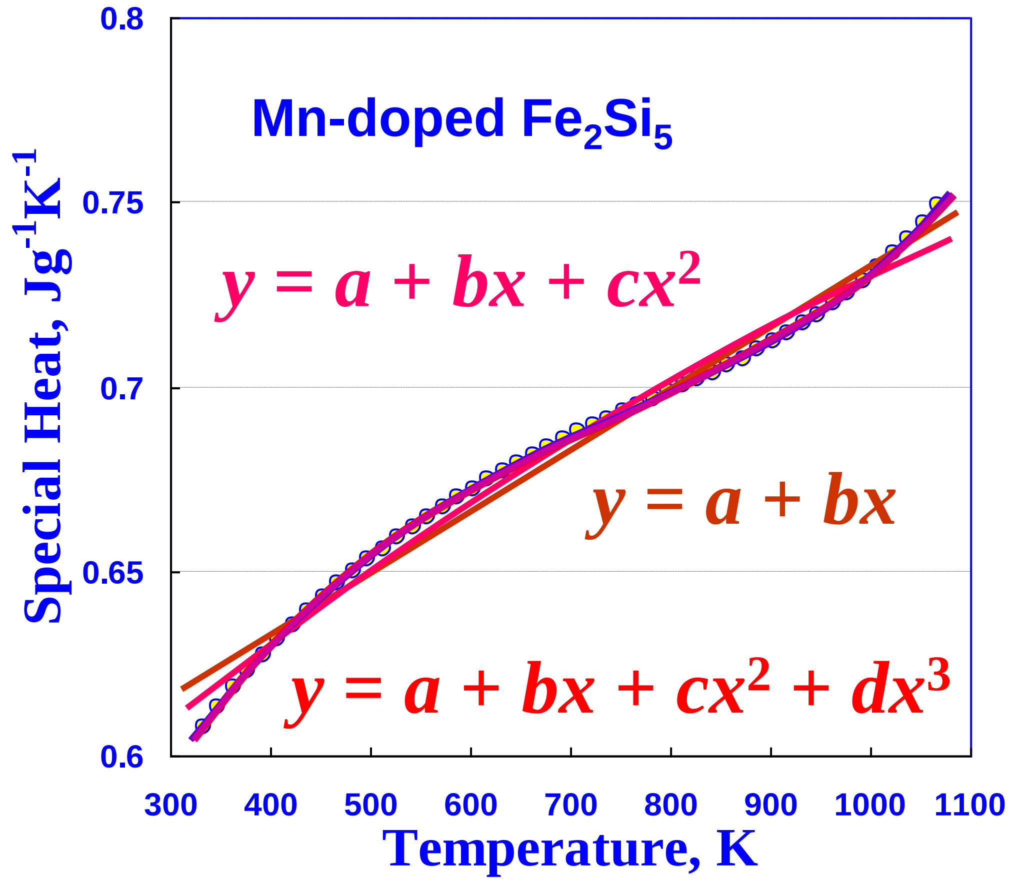
次选常用数理函数曲线

第三选用简单函数曲线

(直线、抛物线、对数)

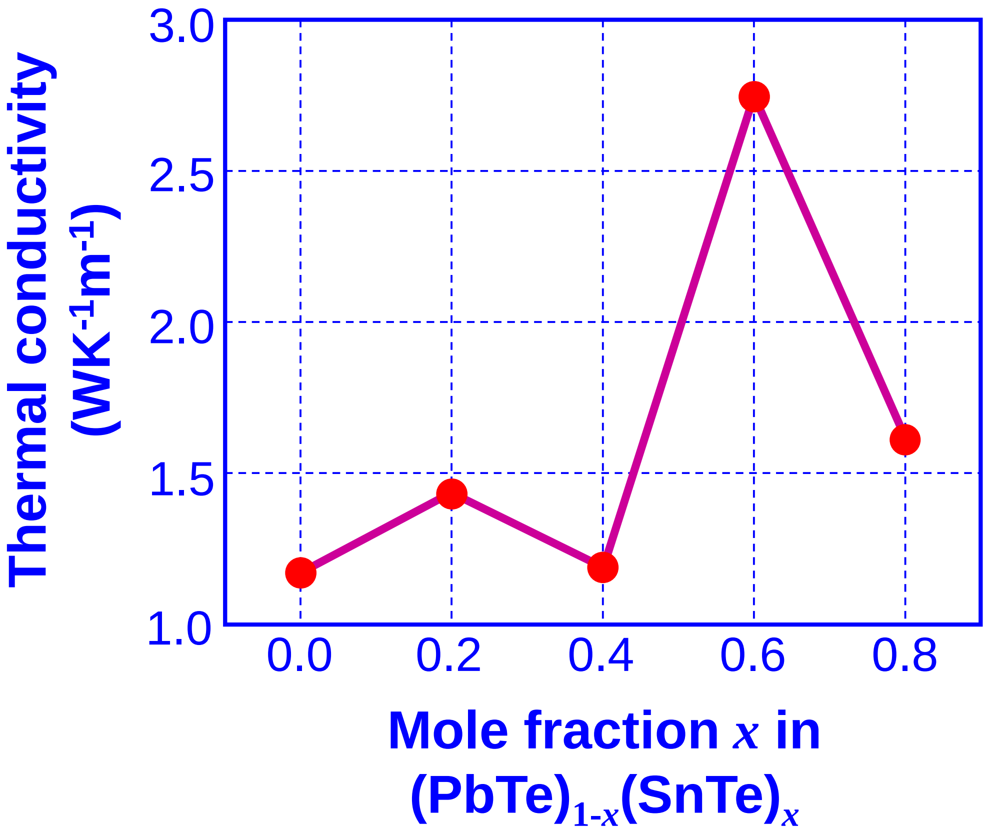
$$C_P = 0.40290 + 9.1456\times10^{-4}T - 1.0622\times10^{-6}T^2 + 4.7551\times10^{-10}T^3$$

$$y = a + bx + cx^2 + dx^3 + ex^4$$



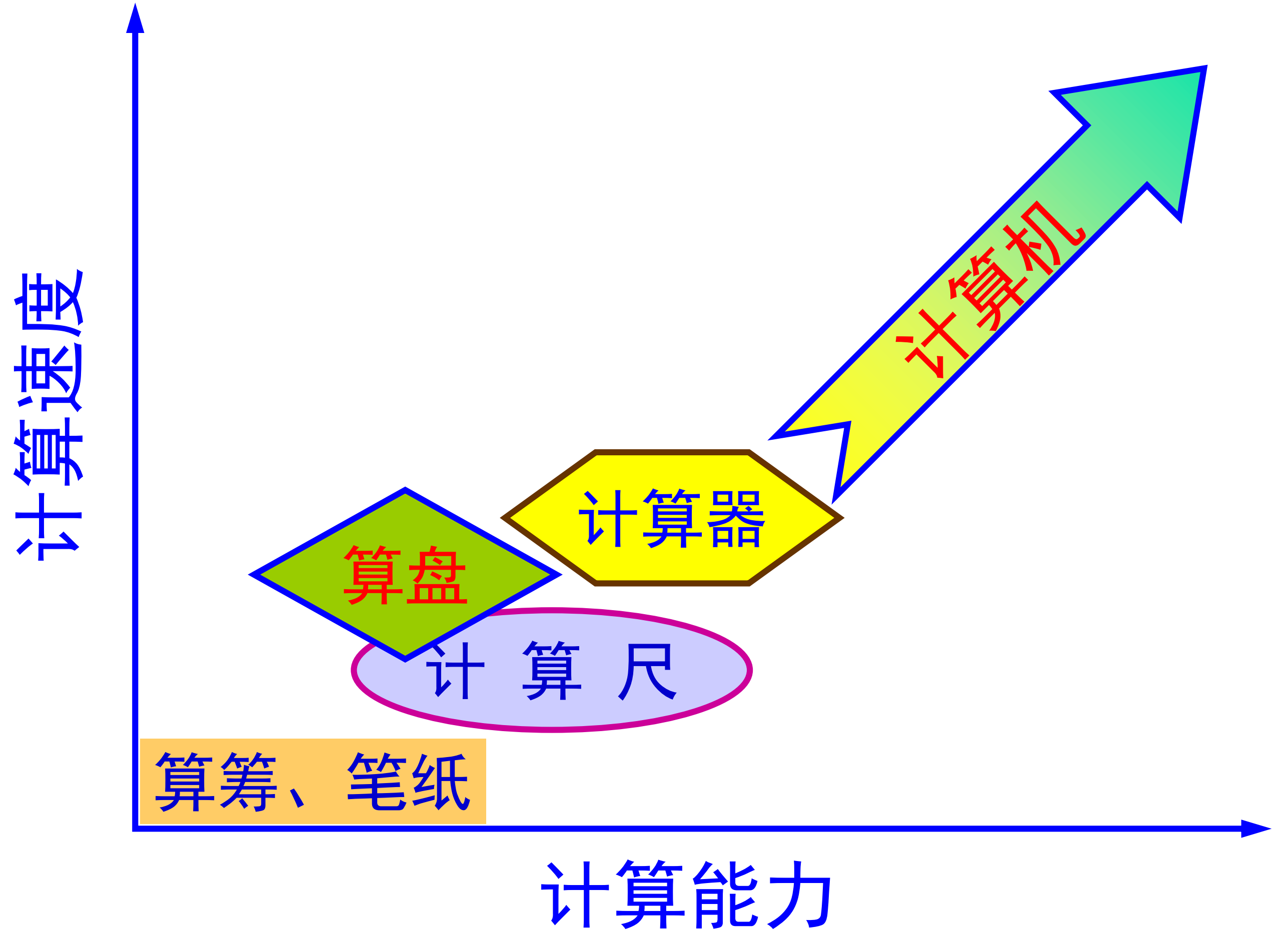
离散数据点之间的连线

首选物理模型拟合曲线
次选常用数理函数曲线
第三选用简单函数曲线
第四直接采用折线连接



忠实反映实验结果
不作任何其他预测

1.5 计算机技术对数值计算的影响



对计算机技术的一些“预测”

1943年，IBM总裁 托马斯•沃森：

我认为也许5台计算机就能满足全世界的需要

1949年，美国《大众机械》：

今后计算机虽很重，但肯定不会超过1.5吨

1957年，数字设备公司创始人 肯•奥尔桑：

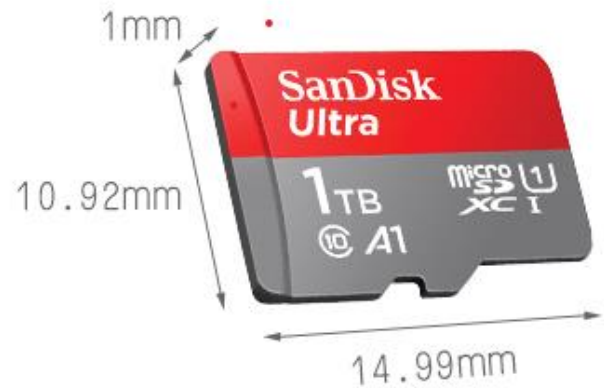
人们在家中使用计算机是完全不必要的

1981年，比尔•盖茨：640 kB 内存对所有人都足够了

2020年，马斯克：人类活在真实世界之中的几率，

可能不到十亿分之一。

数据储存器的发展



(from JD)

计算机技术发展对数值分析的影响

1、算法“加速” 的重要性显著下降

2、储存容量几乎可以不予考虑

3、软件“泛滥”，更需了解原理

具体方法可以急用先学，基本原理必须清晰掌握

4、“数学思想”至关重要

进一步认识“数”：计算机的数，现实中的数

数值处理的目的：提炼物理意义

对“数”的认识

例：　试验测量了三个试样的抗拉强度

结果为：800 MPa，900 MPa，1300 MPa

数值		应用场合
平均数：	1000 MPa	……………写论文
中位数：	900 MPa	……………简单统计
分布范围：	800 ~ 1300 MPa	………检测报告
最大值：	1300 MPa	……………理论家
最小值：	800 MPa	………………工程师

“精确” 和 “模糊”的数

华人思维：模糊

西方思维：精确

请稍等一会。

Just a minute, please.

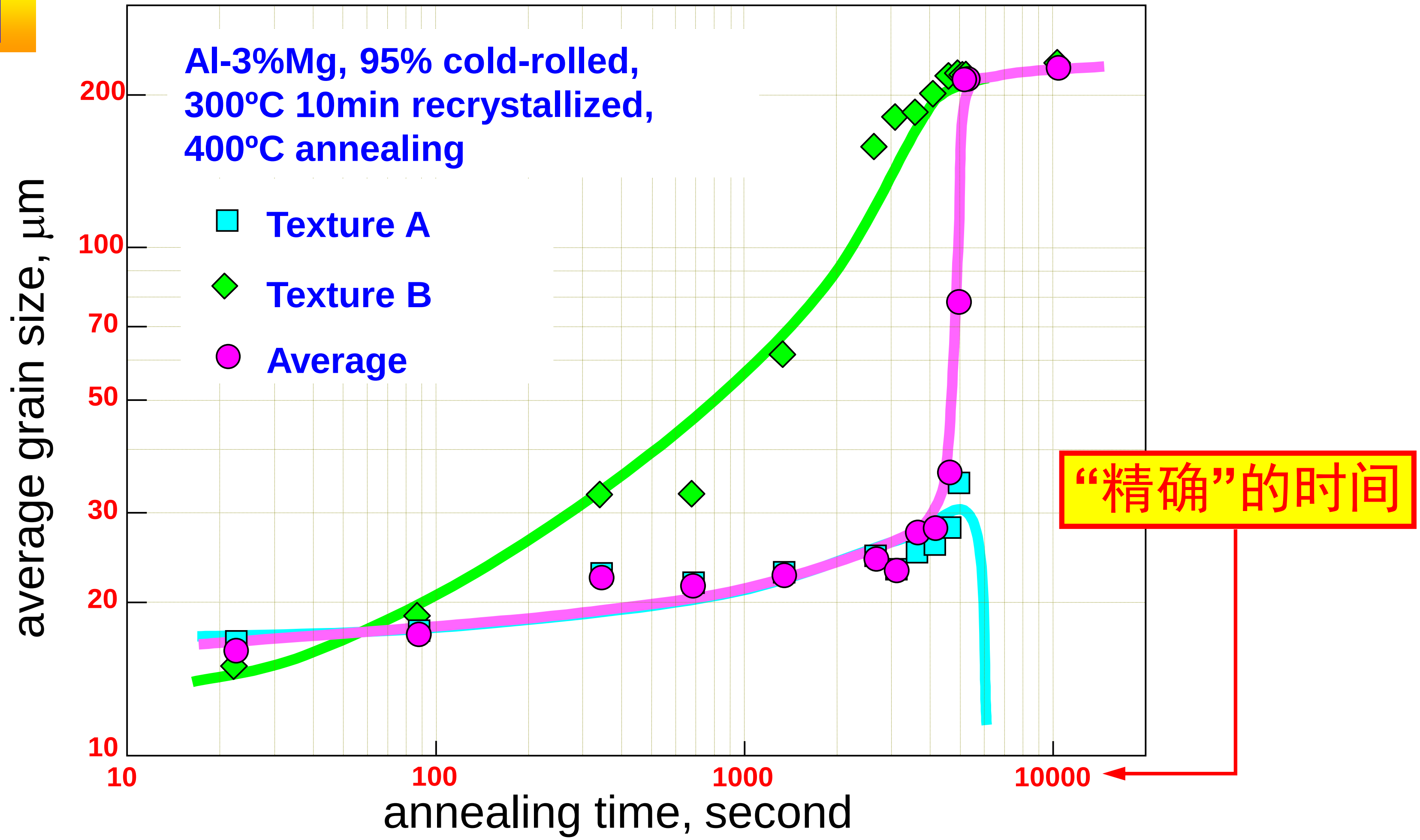
加盐少许。

中火煮片刻。

多功能量杯

微量量匙

“精确”的科学实验



一会儿 片刻 较长 很长 非常长 相当长

小结

基本概念

计算方法

数学思想

课后作业（一）

1. 表中各 x^* 都是对准确值 x 进行四舍五入得到的近似值. 试分别指出其绝对误差限、相对误差限及有效数字位数，并填入表中.

x^*	绝对误差限	相对误差限	有效数字位数
0. 301 2			
30. 12			
30. 120			
30 120			
0.3012×10^5			

2. 若用电表测得一个电阻两端的电压和流过的电流分别为

$$V=110\pm2\ (\mathrm{V}),\quad I=20\pm0.5\ (\mathrm{A}),$$

试由欧姆定律 $R=\dfrac{V}{I}$ 求这个电阻阻值 R 的近似值，并估计所得近似值的绝对误差与相对误差.

补：正方形边长约100cm，边长测量误差多少时才能使其面积误差不超过 1cm^2？