[19]中华人民共和国国家知识产权局

[51]Int. Cl[7]
G06F 15/18
C12Q 1/68

[12]发明专利申请公开说明书

[21]申请号 98811535.2

[43]公开日 2001年1月10日

[11]公开号 CN 1279792A

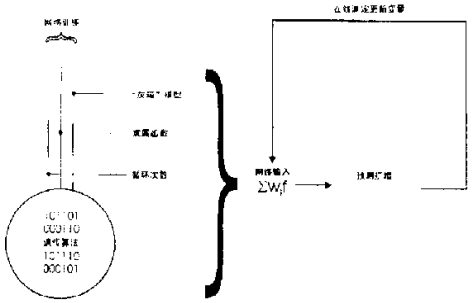
[22]申请日 1998.10.5 [21]申请号 98811535.2
[30]优先权
[32]1997.10.3 [33]GB [31]9720926.6
[86]国际申请 PCT/GB98/02989 1998.10.5
[87]国际公布 WO99/18516 英 1999.4.15
[85]进入国家阶段日期 2000.5.25
[71]申请人 分子感应器有限公司
地址 英国威尔特郡
[72]发明人 约翰·迈克尔·克拉克森
本杰明·戴维·科布
[74]专利代理机构 中科专利商标代理有限责任公司
代理人 姜兆元

权利要求书2页 说明书18页 附图页数15页

[54]发明名称 包括聚合酶链反应(PCR)在内的核酸扩增程序的控制

[57]摘要

一个通过给出变性、退火和延伸事件的隶属关系数值以确定反应中每个事件的 相对基值并利用遗传算法确定完成每一事件所需最适时间来优化用于控制聚合酶链反应的循环条件的系统。



ISSN1008-4274

权 利 要 求 书

1. 一种通过变性、退火和延伸事件建立隶属关系数值的优化用于控制聚合酶链反应的方法，用以确定反应中每一事件的相对基值，和用遗传示方法，确定完成每一事件所需最适时间。

2. 依据权利要求 1 的一种方法，其中，变性、退火和延伸的相对基值通过给出隶属关系数值算出。

3. 依据权利要求 2 的一种方法。其中，隶属关系数值被用来确定在特定时间或通过一系列时间点变性、退火和延伸的相对量。

4. 依据权利要求 1 的一种方法，其中，遗传算法被用来确定变性、退火和延伸的最适时间。

5. 依据权利要求 1 的一种方法，其中，所述方法被用来使 PCR 程序标准化或用来优化 PCR 程序。

6. 依据权利要求 1 的一种方法，其中，所述方法被用来将一个热循环器的程序转移至另一个热循环器。首先，考虑源循环器的热性能，计算变性、退火和延伸的相对基值。然后，在考虑两个循环器性能上的差异的情况下，将它们转移到靶循环器上。

7. 依据权利要求 1 的一种方法，其中，采用在线监控以提供可用来对计算得出的循环时间进行调整的反应表现的反馈信息。

8. 依据任一前述权利要求的一个方法。其中，一个神经网络被用来获得最适循环条件的信息。

9. 依据任一前述权利要求的一个方法。其中，神经网络被用来计算、确认或修改扩增的计算水平。

10. 一种将神经网络用来根据一套给定反应标准预测扩增水平的方法。

11. 依据权利要求 10 的一种方法，其中，给予神经网络的训练输入为正交阵列，而输出代表从一个相关 PCR 得到的扩增水平。

12. 根据权利要求 11 的一种方法，其中，正交阵列被用来减少给

予神经网络的训练输入的数目。

13. 根据任一前述要求的方法，其中，确定的特性是灵敏性设置。

14. 根据任一前述要求的方法，其中，确定的特性是在设置的灵敏度附近可重复的区域。

15. 根据任一前述的方法，其中本方法通过确定预测扩增水平的偏差被用作质量控制程度。

16. 一个优化用于控制聚合酶链反应的循环条件的系统，包括用于建立变性、退火和延伸事件隶属关系数值的处理方法，在反应中，确定每个事件的相对基值；利用遗传算法确定完成每一事件所需最适时间的处理方法。

说　　明　　书

包括聚合酶链反应（PCR）在内的核酸扩增程序的控制

本发明涉及控制聚合酶链反应的循环条件的优化。

优化 PCR 扩增的温度控制需要仔细考虑反应条件。这种反应以及必需反应成分之间的相互作用的复杂特性表明无法简单地采用传统的动力学分析方法来预测最佳循环条件。本方法通过综合采用“灰箱”建模、遗传算法和神经网络对于给定循环条件下的扩增水平进行模型构建和预测来解决上述问题。此法可用来确定温度轨迹的哪一部分对反应具有最大的影响。遗传算法用来建立温度轨迹变化对扩增影响的模型。因而，这些算法可用于确定能够改善反应结果的温度循环。将模型构建与扩增过程的在线监控联系起来，就有可能实现反应的实时优化。这对于诸如基于 PCR 的诊断等质量控制要求高的方法尤为重要。

美国专利第 4683195 号（Mullis et al., Cetus Corporation）公布了一个通过聚合酶链反应（PCR）进行核酸扩增的方法。设计相应于被扩增靶序列两旁区域的、长度通常为 10-40 碱基的短寡聚核苷酸序列。这些引物过量地与靶序列 DNA 混合。再加入合适的缓冲液、氯化镁离子、热稳定聚合酶和自由核苷酸。

热循环过程一般将 DNA 扩增几百万倍。扩增通过温度循环实现。先在 95°C 将靶 DNA 变性，然后通常降温至 40-60°C，使引物与分开的 DNA 链退火。升温至聚合酶的最适温度，通常为 72°C，聚合酶延伸引物、复制靶序列。这一系列事件重复多次（通常为 20-40 次）。在最初的几个循环中，靶序列被复制。在随后的循环中，复制产物被复制，使靶序列呈指数扩增。

由于反应成分之间存在复杂的相互作用，不可能用传统动力学表示法对 PCR 进行数学描述。（参见“A simple procedure for optimizing the polymerase chain reaction (PCR) using modified Taguchi methods”, Cobb

and Clarkson (1994), Nucleic Acids Research. Vol.22, No.18, pp.3801-3805)。已知 Mg^{2+}和三磷酸脱氧核苷会通过改变变性、退火和延伸温度下引物与模板的杂交和解离动力学，影响引发和延伸效率。这些成分还会影响聚合酶识别及延伸上述双链的效率。最佳扩增所需 Mg^{2+}和三磷酸脱氧核苷的浓度在很大程度上取决于靶序列及引物序列，而引物 3'端的核苷酸对于错配延伸的效率有重要影响。某些错配核苷酸组合可能在某些条件下会更有效地扩增。反应中存在过量 Mg^{2+}时可能会造成非特异扩增产物的积累，而如果浓度较低则可能减少产物产量。此外，三磷酸脱氧核苷可结合 Mg^{2+}离子，因而，改变 dNTP 浓度时需要对 $MgCl_2$ 浓度作相应调整。

PCR 的优化通常需要通过对重要的反应参数进行反复的试验-错误调整来完成。这样得到的优化了的反应通常也不是万能的，而是会受到温度轨迹的些微变化和/或反应混合物组成的波动的影响。反应的复杂性以及某种程度上的不确定性表明模型建立并不容易。在已建立的模型中（见'Polymerase chain reaction engineering' Hsu et al., (1997), Biotechnology and Bioengineering, Vol.55, No.2, pp.359-366），重要的反应元素被忽略了。关键在于，由于受热循环器的程序设定方法（即设定相应于变性、延伸和退火步骤的几个固定温度）的影响，目前的模型假定，变性、延伸和退火发生于循环中的固定温度。但这是过分简化，因为这些步骤的速度是温度依赖的，它们发生在一个宽的温度范围之内。

从理论上看，特定模板序列的扩增应遵循指数函数，即，在理想条件下，被扩增的模板的量在每一个反应循环之后应该翻番。但是，扩增的准确性和速率受制于反应成分之间的复杂的相互作用，因此，理论最佳值总是达不到的。在通常条件下，由于在反应靠后的循环中待延伸的双链数目超过了酶活性所能处理的数目，反应产物的积累受到限制。此时，产物积累呈线性增加。使问题更为复杂的是，聚合酶由于长时间暴露于超过 80°C 的温度而热失活。通过仔细考虑退火温度、退火时间和退火斜坡可以优化扩增。通过改变斜坡速度补偿引发率损失，可增加退火温度从而避免非特异引发。这样就可增加靶序列呈指数增加的循环范围。引发率以及产生引发的温度范围取决于游离 Mg^{2+}的量。

由于 Taq 聚合酶过度接触变性高温（一般为≥94℃，1-5 分钟）后会发生失活，在变性时间和斜坡方面作类似的优化将影响扩增（参见'Kinetics of inactivation for thermostable DNA polymerase enzymes', Mohapatra and Hsu, (1996), Biotechnology Techniques, Vol.10, pp.569-572）。尽管通常过量使用聚合酶，但 PCR 反应中一次次变性步骤对聚合酶失活的量有很大影响。另外，这些温度条件会使 DNA 模板发生脱嘌呤（通常在 94℃ 时每分钟每 2kb 一次）。因为在达到设定的变性温度前后变性都在发生（一般在 70℃ 和 90℃ 之间 DNA 变性速度随温度增加），改变斜坡时间可限制暴露于 94℃ 的时间。

聚合酶，如 Taq 的性质已经被很好地研究。它们表现出经典的温度依赖性，在高温时延伸速度逐渐增加，活性达到最大值（通常大约在 70 ℃），然后活性逐渐下降（一般在≥80 ℃ 时）。延伸发生在较宽的温度范围内。根据在这一范围内的总延伸量，可缩短延伸时间。例如，在约 60 ℃ 时，会发生大量的延伸，在这个温度杂交的寡聚核苷酸会立刻被延伸。延伸时间可缩短，在某些情况下可取消。

本发明寻求实现用于控制聚合酶链反应的循环条件的优化。

按照本发明的一个方面可得到在权利要求 1 中描述的优化用于控制聚合酶链反应的循环条件的方法。

首推实施例给出了可实现 PCR 的智能控制的方法。这是通过采用隶属函数建立（反应事件与温度的关系），遗传算法和人工神经网络的一种全新结合对扩增水平进行建模和预测来实现的。隶属函数部分推断决定特定反应扩增程度的各种反应参数并给出清楚的定义。遗传算法用于确定每一温度循环步骤的最适时间。神经网络部分用来增强隶属法则和隶属函数。经过最初的训练，神经网络可在不断从输入信号中得到新东西时更新隶属函数。这一方法可用来精确预测最优反应条件（图 1）。

此法最好用于将一个热循环器的程序移植到另一个热循环器上。首先，根据第一个热循环器的热性能计算变性、退火和延伸的相对基值。然后，考虑两个热循环器性能上的差异，将上述基值移植到第二个热循环器上。

下面通过例子及相伴示意图描述本发明的一个实施例。这些图示包

括：

图 1 为一“智能”热循环器控制的示意图。这一控制利用特定反应事件隶属函数和遗传算法预测设定循环条件下的扩增水平。

图 2 显示采用自学控制过程预测扩增水平。

图 3 为显示模板去向的 S 形隶属函数。

图 4 为采用目标、退火和延伸隶属函数建立 PCR 扩增模型的例子。

图 5(a）为优选遗传算法中关键事件的示意图。

图 5(b)显示图 5(a)中遗传算法的运算。

图 6(a)和图 6(b)分别显示一点和二点交换的例子。

图 7 为人工神经网络的一个设置和一个三层人工神经网络的示意图。

图 8 为 S 形传递函数。S 形增益为 0.25-2.00。

图 9 至图 15 是与优选方法相关的结果。

近来已有实时监控 PCR 过程的新技术（例如：5'端荧光核酸酶化学—PE Applied Biosystem 和溴化乙锭荧光（见"Kinetic PCR analysis real-time monitoring of DNA amplification reactions" Higuchi et al., (1993), Biotechnology. Vol.11, pp.1026-1030）。这些方法可对反应过程中形成的产物进行定量，而实时监控的主要好处产生于可精确预测和保持最佳扩增的算法。本申请中介绍的方法可与产物检测系统结合，为 PCR 提供实时动态监控，随反应进行连续更新循环条件，保持最佳运行。

通常需要仔细考虑 PCR 条件以优化扩增水平。但由于聚合酶链反应的复杂性，不能简单地运用传统动力学方法预测最佳条件。加之，反应成分之间还存在复杂的相互作用。首选实施例寻求通过计算循环途径中整个温度范围内变性、退火和延伸的量预测扩增水平来克服一直与 PCR 优化相关的问题。采用将隶属函数运用于上述每一事件的“灰箱”建模、遗传算法和人工神经网络的全新结合，可以预测扩增水平。对神经网络产生的权重的分析可用来在循环事件的每一步所需时间上定义反应最佳点。这个过程可描述如下：

i 采用灰箱建模定义最适退火、延伸和变性温度和这些事件发生的温度范围。

ii 将隶属函数运用于每一事件，从而预测给定循环的扩增水平。

iii 运用遗传算法确定每一阶段的最适时间。

iv 可运用神经网络确认和/或更改预测的扩增水平。

v 实时监控可用来进一步完善这一过程。

这一方法为适用于 PCR 的控制软件提供了基础。这类软件可运用于标准热循环器控制。这一方法还首次介绍了优化 PCR 的智能控制过程。

定义成员函数

优选方法的基础包括将 PCR 的重要元素（变性、退火和延伸）转变为一系列隶属函数。首先，通过对反应进行“灰箱”建模建立各种影响反应特性的参数的隶属函数系列和法则。参数指影响模板变性、引物退火和引物/模板双链延伸的因素。用一个规则库将这些事件与扩增循环中特定温度联系起来。每一反应变量的明确数值被用来预测单个或多个循环的扩增水平。采用隶属函数能够建立 PCR 的动态模型，考虑整个温度循环中，而不象传统建模方法中只是特定阶段的，各种过程的速度的变化。

可以认为，PCR 包含三个主要事件，即双链模板的变性，引物退火至单链变性模板上以及这些双链的聚合。这些过程进行的速度依赖于温度。反应中各种成分之间的相互作用也影响速度和最佳值。建立这些过程的“灰箱”模型使这些事件的隶属函数得以与特定温度相关联（图 2）。遗传算法可用来调整分配给每个阶段的时间（t_1-t_6）以优化扩增，限制组分相互作用、失活、脱嘌呤等的影响。神经网络部分被用来了解开始时可用的时间以减少计算阶段最佳值的时间。

变性隶属函数

DNA 模板的变性可用 S 形融解曲线表示。模板的退火温度定义为一半 DNA 变性时的温度。70°C 以下几乎不发生变性。在此温度之上变性速度迅速增加。酶失活和模板脱嘌呤速度在高温时也增加。变性主要发生在循环中的一个温度范围内，即，在温度上升至设定变性温度，从这一温度下降和在这一温度时。计算在达到设定变性温度前后以及在变性温度时的变性速度可以大大减少设定变性时间。这就将酶失活和脱嘌呤

量降至最低。因此，扩增效率得以提高。

双链模板的叠加而成的、氢键连接的碱基对之间的具有协同性的相互作用在融解时逐步被破坏直至两条链完全分开。模板变性的隶属函数（$M_{denature}$） 因而可以用 S 形曲线定义（图 3），增益为 $q_{denature}$，模板分子的预测 T_m 和最适变性温度定义的中点为 $S(X)_{denature}$。增益 $q_{denature}$ 和中点 $S(X)_{denature}$ 用作神经网络部分的修正量。归纳出的曲线还可以用来定义变性隶属关系（例如，布尔型,梯形，三角形和动态的）。通常，含 AT 和 GC 碱基对的天然 DNA 的 T_m 值约为 72℃，多聚 AT 模板 T_m 值约为 60℃，和多聚 GC 模板的 T_m 值约为 90℃。

退火隶属函数

退火隶属关系定义引物与模板的杂交速度。开始时，引发温度 T_p（或 T_m-5℃）用来定义取决于引物长度的最适杂交。通常，T_p 和 T_m 由下列方程式之一得到：

- $T_p = [22 + 1.46 \cdot (2 \cdot GC + AT)]$
- $T_m = 81.5 + 16.6 \cdot \{(\log_{10}[J^+]) + 0.41 \cdot (\%G+C) - (600/l) - 0.63(\%FA)\}$
- $T_m = 81.5 + 16.6 \cdot \{(\log_{10}[J^+]) + 0.41 \cdot (\%G+C) - 675/l\}$
- $T_m = [(A+T) \cdot 2°C] + [(G + C) \cdot 4°C]$

其中，J^+为单价阴离子浓度，l 为寡聚核苷酸长度，FA 为甲酰胺。杂交隶属函数（M_{anneal}）为 S 形曲线(图 3).其最大值由引物 T_p 值或 T_m-5 ℃ 决定。归纳出的曲线还可以用来定义隶属关系（例如，布尔型，梯形，三角形和动态的）。发生退火的温度范围取决于反应混合物中自由镁离子的浓度。中点 $S(X)_{anneal}$ 和增益 q_{anneal} 定义发生退火的温度范围。两者与反应中自由 Mg^{2+}浓度紧密相关。增益 q_{anneal} 和中点 $S(X)_{anneal}$ 用作神经网络部分的修正量。

延伸隶属函数

别人先前已经详细描述了 Taq 活性对温度的依赖性（参见 "Deoxyribonucleic acid polymerase from the extreme thermophile THERMUS aquaticus", Chien et al., (1976), Journal of Bacteriology. Vol.127, No.3, pp.1550-1557）。双链聚合的隶属函数由一最大值由最适温度定义的曲线表示。最适温度将取决于所用的聚合酶。一般在 70 °C 左右。归纳出的曲线也可以用来定义隶属关系（即布尔型，梯形，三角形和钟形的）。

图 4 为一单个 PCR 循环时间为 t 秒时样品管温度的反应隶属关系 $M_{denature}$, M_{anneal} 和 $M_{extension}$ 的数值。调整这些循环事件之间或事件之中的时间 t 可优化所能得到的扩增量。由于优化过程以这些事件的减少为先决条件，因此在优化过程中降低扩增的因素（例如，脱嘌呤、Taq 聚合酶失活等）会被减少。神经网络算法用来确定这些事件中哪些对反应结果影响最大。

采用遗传算法确定最适轨迹时间

由于改变变性、退火和延伸任一设定时间都会改变扩增水平，优化反应的方法不是一目了然的，也不是计算量很大的。这里介绍的方法非常独特地利用遗传算法克服这些困难。保存一系列可能的问题解决方案，并根据进化原则（即选择、突变和/或重组（交换））不断更新系列。重组从这个系列（亲本）中选出方案对，并通过结合亲本元素产生一系列新方案（子代），将新方案插入新的方案系列。选择的原则要求“好”的方案优先于“坏”的方案被选中。这可以通过定义适合度函数做到。适合度函数根据每个方案的“好”的程度给它一个数值。选择会确保只有适合度最好的方案（染色体）在将来的系列中增殖。突变被用来修改系列中的一些方案而不影响其他方案。因此，遗传算法寻求在一起时会形成好的方案的基因组合（共适应基因）。

遗传算法的一般性定义并不存在。这些算法可以从 Holland 的工作中抽象出来（参见“Adaptation in natural and artificial systems”. Holland,

(1975). The University of Michigan Press; Ann Arbor)。本方法中使用的遗传算法的示意图见图 6。它有下列步骤：

i　启动 染色体的最初系列（通常 n=25…100）可随机建立或者通过将下文介绍的人工神经网络部分产生的输入染色体“拆装”来建立。这就缩短了算法用来寻求最适值的时间。

ii　评估 评估每个染色体的适合度。采用前已说明的隶属函数以及单个染色体给定的循环时间计算扩增量。然后，给出这些染色体的适合度函数。这些函数会给每个染色体的表现进行数值编码。

iii　发掘 将具有最高适合度分值（即在最短的每步耗时情况下的最高预测扩增）的染色体半随机地、一次或多次置于配对子集中。从系列中随机抽取两个染色体。将具有最高适合度分值的染色体置于配对子集中。将两个染色体放回到系列中。重复选择过程直到配对子集排满。这就确保低适合度染色体从配对集合中剔除。被选作亲本的机会与染色体的标准适合度呈正相关。这意味着好的染色体一般将产生较多的后代。但是，这一过程的随机性意味着较差的方案偶尔也会产生后代。增加优越函数也是有好处的。这样，任一亲代中最好的方案不加改动地被复制到子代中，所有其他子代均按常规得到。

iv　探索 使用重组和突变算符修改后代染色体中的染色体（图 5）。随机选择配对子集中两个染色体并交配。交配概率为一可控制的功能，一般用较高值（≈0.90）。重组算符用来在两个亲本染色体之间进行基因交换，产生两个子代。图 5 显示一个三个个体的系列。每一个体有一适合度函数 F。基于这些适合度，选择过程给予第一个个体 0 个拷贝，第二个个体 2 个拷贝，第三个个体 1 个拷贝。选择后，按概率使用遗传算符，第一个个体的第一个字符由 1 突变为 0，其余二个个体经交叉形成新个体（修改自 Forrest (1993). Science, Vol.261, pp.872-878)。重组可以是一点和/或二点交叉（图 6)。使用一点交叉时，在染色体上选择一交叉点。到那一点为止的两个亲本的基因发生交换。发生二点交叉时，选择二个交叉点。两点之间的基因发生交换。其他交换模式，如部分图解交换(PMX)，顺序交换(OX)和循环交换(CX)也可采用。所有子代产生于两个亲本的交换。二个子代个体同时产生。子代在下一代中取代亲本。单基因突变是

另一可控功能，通常用低突变率（≈0.001），定义为（$N*L^{1/2}$）$^{-1}$。其中，N 是系列大小，L 是染色体长度。

v 这一系列的事件重复一定代数直到发生收敛。拥挤取代可用来减少过早收敛。一个子代个体与大量现有亲本个体比较，并取代最象自己的一个亲代个体。拥挤取代本质上就是相似个体之间的取代。这就使亚系列能够检查遗传搜寻空间的不同部分。这对于具有大量的、分散适合度极值的多形搜寻空间特别有用。

遗传算法的基本结构可以归纳如下：

建立系列

评估染色体

计算标准化适合度

关于系列的输出统计学

对每一个下一代

（产生和评估侯选置换者

以侯选者取代系列中的成员

重新评估未改变成员

适合度标准化

关于系列的输出统计学）

根据循环中特定阶段的最短/最长时间和所需的时间分辨率（表 1），循环轨迹中的每一个时间（基因，t1…t6）转换为位 $I=\{0,1\}^1$。这些位组合起来形成代表单个循环总的时间轨迹的位串（染色体）。首先，根据给每个阶段的最小和最大时间，建立一个随机系列位串。可使用经过根据变性、退火和延伸步骤的隶属函数训练已能够设置初始时间的神经网络（注：传统的二进制编码有一个缺点，即在某些情况下，为将数值改变 1，所有位都必须改变。这就使通过突变使一个已经接近最佳值的个体更接近最适值较难。由于采用葛莱码时，任何数值增加或减少 1 总是单位变化，因而较为可取。）

表 1 编码 t6，即热循环器由退火（72°C）加热至变性（94°C）所需时间的位数据。最大允许时间设为 120 秒（每秒 5°C 过渡）

解离	位大小	位数据	Σ(t1...t6)
1 秒	7 位	1111 000	42 位
2 秒	6 位	1111 00	36 位
5 秒	5 位	1100 0	30 位
10 秒	4 位	1100	24 位

循环中每个事件的最短时间是热循环器性能限定的。一般而言，最低斜坡速度设为 0.5 °C/秒到 1 °C/秒之间。最高斜坡速度通常无需设得比 5 °C/秒高。因此，循环步骤的最长时间取决于斜坡速度和相继步骤之间的温差。

扩增采用由系列中每一染色体决定的时间轨迹和已经介绍的隶属函数建立模型。适合度打分是根据在可能的最短时间内保持对极端温度最少暴露情况下获得的最大扩增。极端温度可能增加引发错误、失活和脱嘌呤。在某些情况下，定义其他适合度特征可能是有益的（例如，对于 RAPD、差异显示等，可能需要增加引发错误）。在系列中的所有个体都被评估后，它们的适合度用作进行发掘和探索、选择用于下一轮选择的染色体的基础。经过一定数目的重复，最适温度轨迹在染色体系列中的出现频率会很高。

神经网络选择

本申请使用前馈神经网络。它具有一个输入层，一或多个隐藏层和一个输出层(图 7)。每个节点含代表外部信号或其他节点输出的一系列加权输入 Wi。隐藏节点的数目为可变参数。加权输入之和采用非线性 S 形变换函数变换。

$$f(x)=\frac{1}{1+e^{-x/\theta}}$$

其中，f(x)的范围是 0-1，x 是输入的加权和，q 是增益。改变 q 会改变曲线形状。Q 值小则 S 形函数的斜率较陡（例如，q=1.0）。相反，q 值大则曲线斜度较缓（例如,q=2.0）。输入表示温度控制事件的瞬变时间和这些事件的温度范围（图 8）。每一个变量有一个输入节点。输入节点将加权输入信号传递到隐藏层中的节点。输入层中的节点 i 和隐藏层中的节点 j 之间的连接用加权因子 Wij 来表示。因此，对于隐藏层中的每一个 j 节点都存在一个权重向量 Wj。这些权重在训练过程中被改变。每层还有一个可解决数据中非零偏差的偏向输入。偏向值开始时总是设成能够调零。权重矢量中包括将偏向与相应层相接的一项。这项权重也可在训练时改变。可以运用其他函数（正切、正铉、余玄、线性等）。

在初期训练程序中，向网络重复提供一系列输入模式及其相应预期输出。同时，一边调节权重。继续这一过程直至形成预期实测输出之间的所需程度的感觉。可以采用不同的学习算法。但反传是目前的优选算法。预期输出的误差反传通过网络，利用广义增量法则确定对权重的调整（参见"Parallel distributed processing explorations in the microstructure of cognition". Part 1. Rumelhart and McClelland, (1986). MIT Press; Cambridge, MA）。输出层项由下列公式给出。

$$\delta_{pk}=(t_{pk}-0_{pk})0_{pk}(1-0_{pk})$$

其中，d_{pk}为输出节点 k 的观察 p 的误差项，t_{pk}是观察 p 的预期输出，o_{pk}是实际节点输出，$o_{pk}(1-o_{pk})$为 S 形函数的导数。隐藏层的节点 j 的误差项为 S 形函数导数乘以输出误差项与输出层权重之积的和。

$$\delta_{pj}=0_{pj}(1-0_{pj})\sum_{k-1}^{k}\delta_{pk}w_{kj}$$

得自输出和隐藏层的误差项通过调节它们相应层的权重反传通过网络。权重调节，或δ权重，可按下列公式计算。

$$\Delta w_{ji}(n) = \eta \delta_{pj} 0_{pj} + \alpha \Delta w_{ji}(n-1)$$

其中，ΔWij 为隐藏层中节点 j 与输入层中节点 i 之间的权重的变化。η 为学习速度，d_{pj} 为隐藏层节点 j 观察 p 的误差项，o_{pi} 为输入层节点 i 观察 p 的测得输出，a 是动量，n 和 n-1 分别为当前和过去的迭代。整套 p 训练观察在重复次数 n 超过 p 时要重复表示。相似的方法可用来调节连接节点的隐藏层与下一个隐藏层之间的以及末端隐藏层与输出层之间的权重。训练前，所有权重均用随机数值。

神经网络算法是用来联系隶属函数与定义每一扩增阶段时间轨迹的。训练后，这个程序可用来非常精确地预测扩增水平。比较预测的扩增水平与实时监控可帮助进一步优化反应。

B.D. Cobb 和 J.M. Clarkson（1994）的方法提供了评估各种反应要素对反应扩增水平的影响的快捷方法。这可用于训练此法的神经网络部分。神经网络从反应中获得关键信息来预测扩增水平。一个采用 5，5，3，1 格式、具有二个隐藏层的、五道输入的网络被用来从退火温度、模板、引物、dNTP 和 Mg^{2+}浓度等数据预测扩增。利用表 2 条件进行了 27 个训练反应。

表 2 用来训练一个含五个输入节点、二个分别具有五个和三个节点的隐藏层和一个信号输出层的四层神经网络的反应条件/扩增水平

反应	温度	引物	模板	dNTPs(mM)	Mg^{2+}(mM)	扩增
1	45	10	50	0.1	1	2
2	45	10	100	0.2	2	2
3	45	10	150	0.3	3	5
4	45	20	50	0.2	3	5
5	45	20	100	0.3	1	0
6	45	20	150	0.1	2	0
7	45	30	50	0.3	2	0
8	45	30	100	0.1	3	0
9	45	30	150	0.2	1	0
10	50	10	50	0.1	1	2
11	50	10	100	0.2	2	3
12	50	10	150	0.3	3	4
13	50	20	50	0.2	3	1
14	50	20	100	0.3	1	3
15	50	20	150	0.1	2	4
16	50	30	50	0.3	2	5
17	50	30	100	0.1	3	4
18	50	30	150	0.2	1	3
19	55	10	50	0.1	1	0
20	55	10	100	0.2	2	1
21	55	10	150	0.3	3	5
22	55	20	50	0.2	3	5
23	55	20	100	0.3	1	0
24	55	20	150	0.1	2	2
25	55	30	50	0.3	2	0
26	55	30	100	0.1	3	3
27	55	30	150	0.2	1	0

从每个反应获得的扩增量按 0 到 5 打分。这些数据被用来通过反传和 S 形转换函数训练一个神经网络。当然，这一过程并不局限于采用这些转换函数和学习算法。图 9-15 显示训练过程中神经网络的输出以及每

次输入是怎样相互关联的。从这些数据可见，反应最适温度显然在 50℃左右。表 3-5 介绍了网络信息（注：采用这些数据时，遗传算法可用来试验许多输入信号并选择最大输出，从而定义反应最适值。）

表 3 网络训练信息

网络名称：	无名称
层数：	4
输入层：	
布：	5
转换函数：	线性
隐藏层 1：	
节点：	5
转换函数	S 型
隐藏层 2：	
节点：	3
转换函数	S 型
输出层：	
节点：	1
转换函数	S 型
连接：	完全
训练信息：	
重复：	775
训练误差：	0.002895
学习速率：	0.111424
动量因素：	0.800000
快速扩增系数：	0.000000
输入节点：	C:/qnet97t/samples/PCR Input.txt
输入起始柱	1
标传输入：	是
输出节点：	C:/
输出起始柱：	1
标准输出：	是
训练模型：	27
测试模型：	0

表 4 一个四层神经网络 PCR 扩增训练目标和预测输出

反应	目标	预测
1	2.000000	2.008679
2	2.000000	1.998197
3	5.000000	4.999532
4	5.000000	4.993848
5	0.000000	-0.065793
6	0.000000	-0.005967
7	0.000000	0.007813
8	0.000000	-0.000233
9	0.000000	0.048163
10	2.000000	1.996414
11	3.000000	3.010950
12	4.000000	3.986164
13	1.000000	1.012590
14	3.000000	2.998190
15	4.000000	3.986575
16	5.000000	4.986286
17	4.000000	4.007507
18	3.000000	3.012231
19	0.000000	-0.060089
20	1.000000	1.000848
21	5.000000	5.024660
22	5.000000	4.97064
23	0.000000	0.000820
24	2.000000	1.995387
25	0.000000	0.066374
26	3.000000	2.999384
27	0.000000	-0.000950

表 5 网络权重及调整增量

网络权重和调整增量

网络名称：无名称

重复：10000

层	节	连接	权重	权重增量
2	1	1	1.15091	0.000014
2	1	2	0.86859	0.000007
2	1	3	1.72548	0.000003
2	1	4	-2.79642	0.000000
2	1	5	-6.97007	-0.000008
2	1	1	1.77795	-0.000003
2	2	2	-0.03676	-0.000008
2	2	3	4.06031	-0.000007
2	2	4	-0.75059	-0.000002
2	2	5	-4.99932	0.000007
2	3	1	-2.30613	-0.000104
2	3	2	-2.38492	-0.000003
2	3	3	-0.72564	-0.000005
2	3	4	0.35246	0.000009
2	3	5	4.50212	0.000009
2	4	1	-10.99939	0.000005
2	4	2	4.12186	-0.000005
2	4	3	-0.27209	0.000008
2	4	4	0.64145	-0.000016
2	4	5	-0.63613	-0.000005
2	5	1	-7.46985	-0.000008
2	5	2	-1.74181	0.000004
2	5	3	2.85897	-0.000001
2	5	4	-0.11206	0.000003
2	5	5	3.16325	0.000007
3	1	1	-1.70294	-0.000002
3	1	2	2.71630	-0.000002
3	1	3	-1.11906	-0.000002
3	1	4	5.30498	0.000015
3	1	5	-3.25660	-0.000011
3	2	1	5.48513	0.000007
3	2	2	4.11580	0.000003
3	2	3	-0.09909	-0.000007
3	2	4	6.50361	0.000002
3	2	5	-7.38153	-0.000009
3	3	1	0.12320	0.000000
3	3	2	1.73109	0.000007
3	3	3	-4.40772	0.000000
3	3	4	-0.02973	-0.000009
3	3	5	-1.14767	-0.000002
4	1	1	6.41485	0.000013
4	1	2	-7.82154	-0.000013
4	1	3	3.38027	-0.000002

应用

目前，PCR 技术面临的主要挑战是发展能够控制温度循环轨迹优化扩增水平的“智能”仪器。这需要大量技术注入以定义循环条件。最佳值的发现需要通过反复的试验-失败实验。从实验室时间、操作者时间和物质消耗考虑，这是昂贵的。最适条件常找不到。而使用次最适条件时，难以解释和分析扩增结果。

这里介绍的方法可用来通过对循环条件的智能控制优化热循环条件。它利用这样一个事实，即变性、退火和延伸不局限于固定的温度，而发生于循环中的一个温度范围内。建立这些事件的全循环的模型使每个事件的时间得以优化。作为优化的结果，错误引发、失活和脱嘌呤事件也因减少了接触促进这些事件的条件而减少。这一方法可以通过考虑加热块与样品温度之差而进一步改善。这两者之间有明显滞后。这种滞后对扩增水平有很大影响。

用反应混合物的细节（如 Mg^{2+}浓度，dNTPs,引物序列，模板数据等）建立隶属函数。用反应数据和预测模板大小建立反应模型、确定扩增水平。用遗传算法优化循环的时间形态。这就不需要操作者编程，提供了程序标准化的框架。

将扩增的在线监控与本法相结合，可以实时、动态地维持最适扩增条件。这在诊断方面特别有用。在诊断应用上，保证扩增的有效是质量控制的关键因素。比较预期的和实际的扩增可以确定反应在哪里出现问题。这些信息可用来更新循环条件以维持最适扩增或提醒使用者扩增出了问题。对于 NAA 技术未来的挑战包括将这项技术从医院实验室的分析者下放，从而实现实地应用（例如，在地方卫生诊所应用）。医院实验室由于实验室空间以及雇佣技术人员而有较高的附加开销。上述技术下放有很多问题。不良操作条件、样品制备上的差别、操作者的失误均会影响扩增及数据解释。对能补偿操作环境变化的稳定的扩增条件和质控程序的需要将使基于 PCR 的诊断更快地被接受。

作为一种定性试验，重要的性能特点应被了解，并针对每一应用进行确定。样品类型和诊所条件会影响 NAA 试验结果的解释。由于临床样品中存在抑制剂，操作环境的改变或操作者的错误，经常出现假阴性

的 PCR 结果。假阳性结果可能由于试剂被靶序列污染而出现。灵敏度设置（例如，对基因拷贝数、退火温度的灵敏度）和在设置的灵敏度附近重复性差的区域可根据 PCR 性能特点弄确切。这些特点以及样品类型和测试的临床目的为考虑质控提供了框架。本申请提供了维持最适扩增、迅速确定决定特定扩增临界值的因素的基础。

采用我们的方法优化的反应本身具有较好的可靠性。因此，这个过程可用作标准优化程序的基础。对于诸如循环仪器性能变化等变量的可靠性会增加特异性和灵敏度。此外，从本程序中得到的大量信息为实现反应表现的严格质控提供了基础。从有限数量的反应中即可获得关于热循环器运行表现、反应混合物中必需成分、操作环境和提供测试核酸材料的量的变化的重要信息。重要的一点是，首选方法提供了一种“内置”在不同条件下操作的能力的方法。这种方法还通过连续监控分析程序的表现反馈优化随后的分析。

本申请对其享有优先权的英国专利申请第 9720926.6 号中公布的内容以及伴随本申请的摘要中公布的内容作为参考纳入本文。

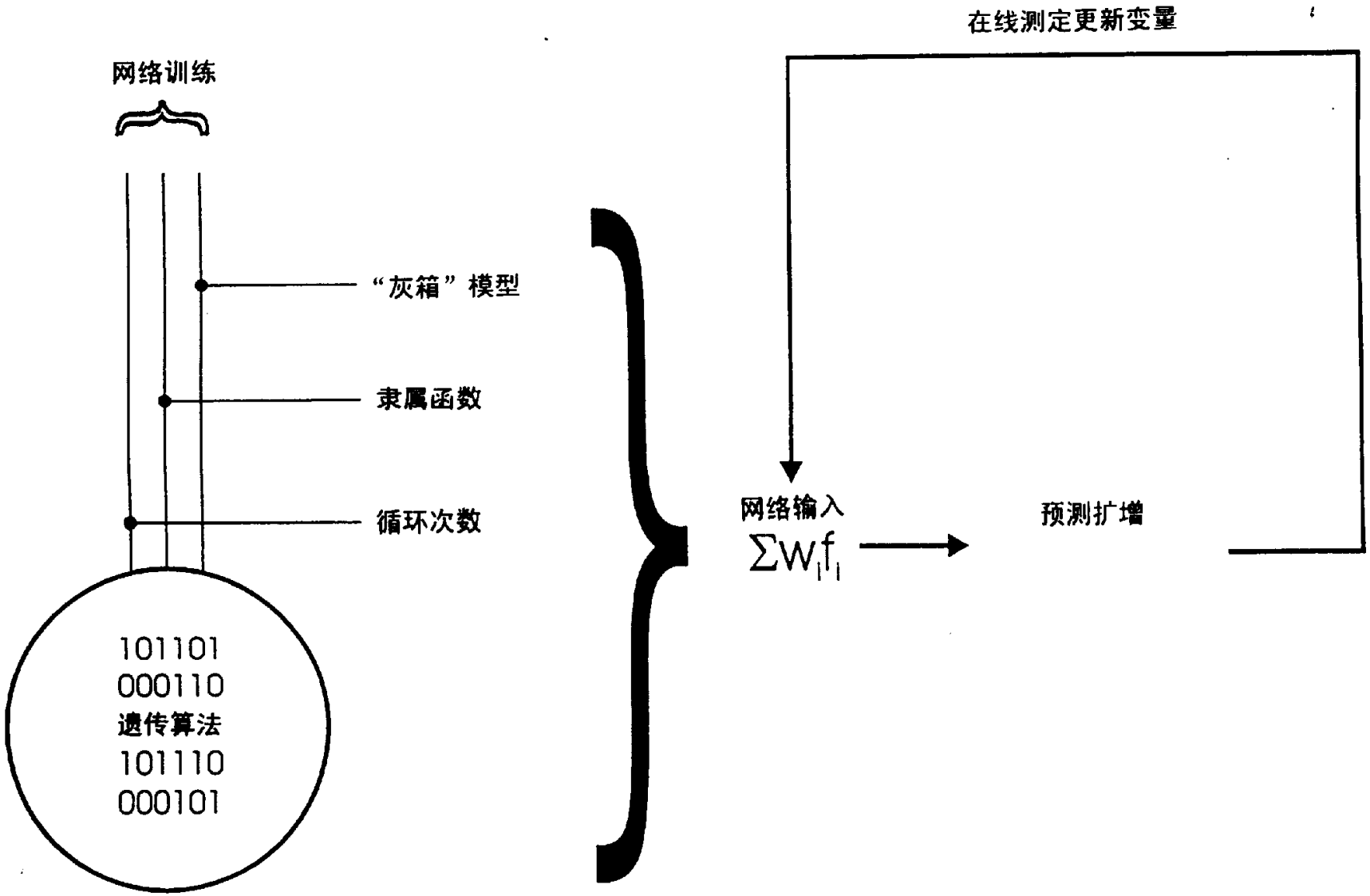

网络训练
“灰箱”模型
隶属函数
循环次数
101101
000110
遗传算法
101110
000101
在线测定更新变量
网络输入
$\Sigma w_i f_i$
预测扩增

图 1

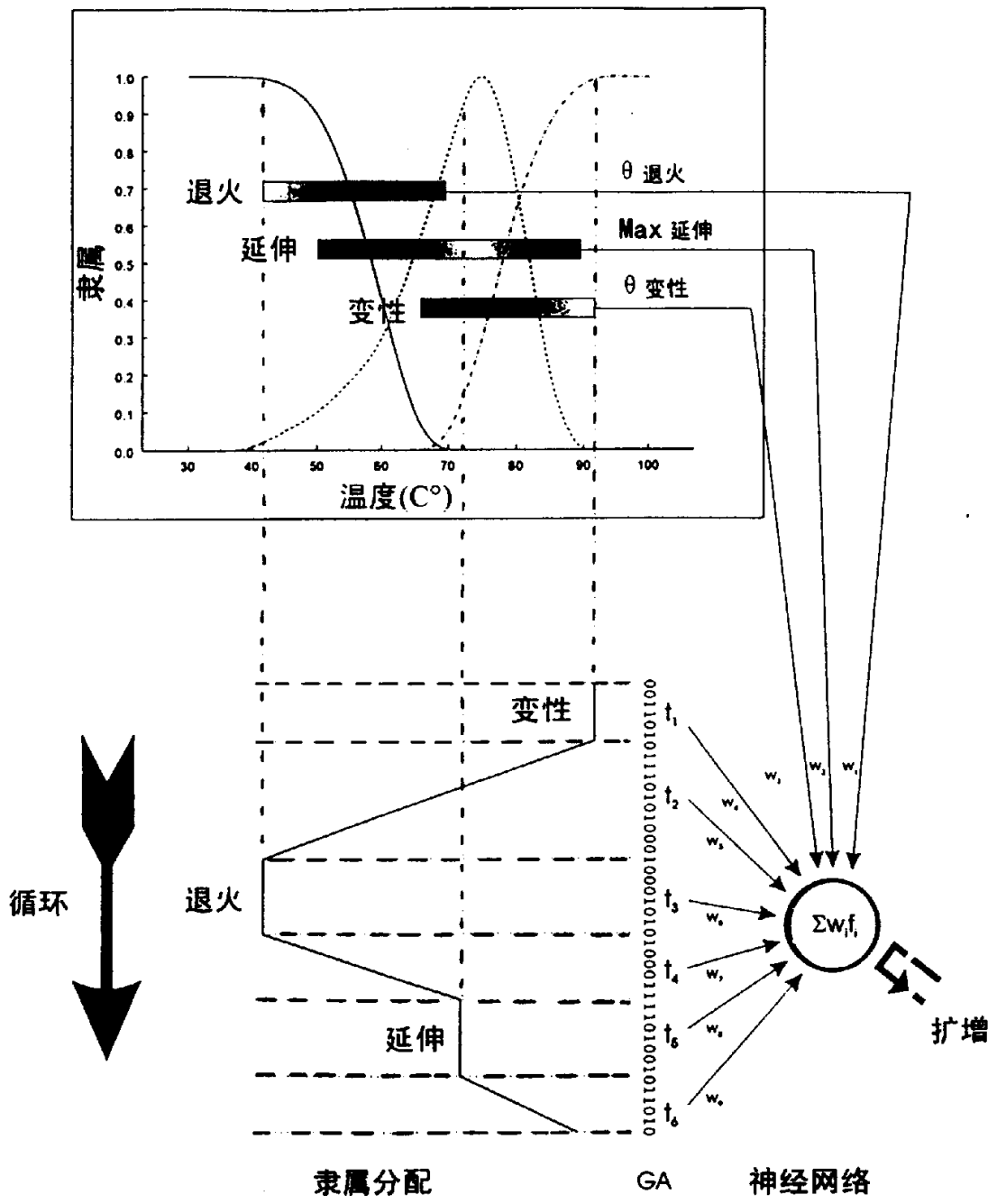


图 2

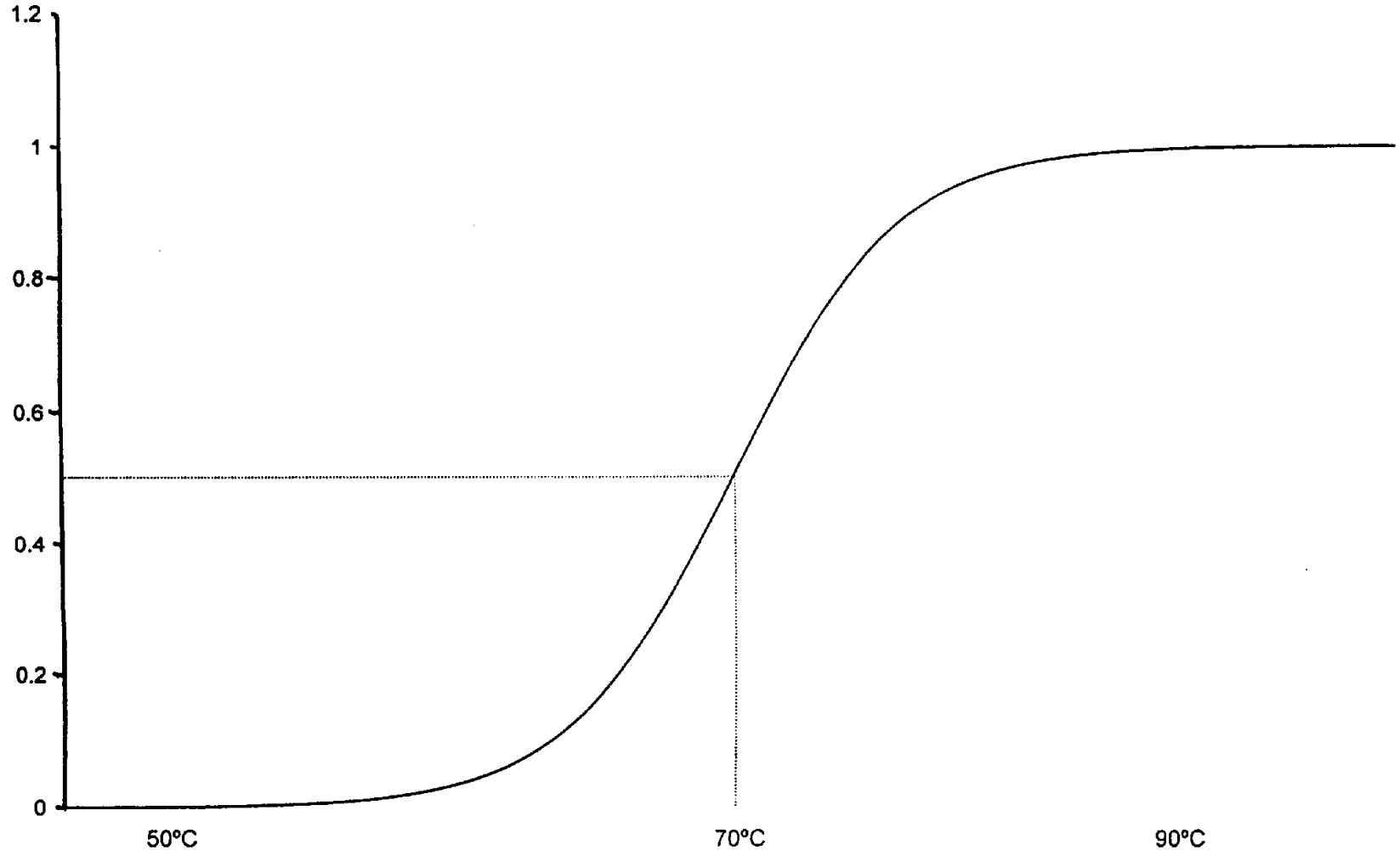
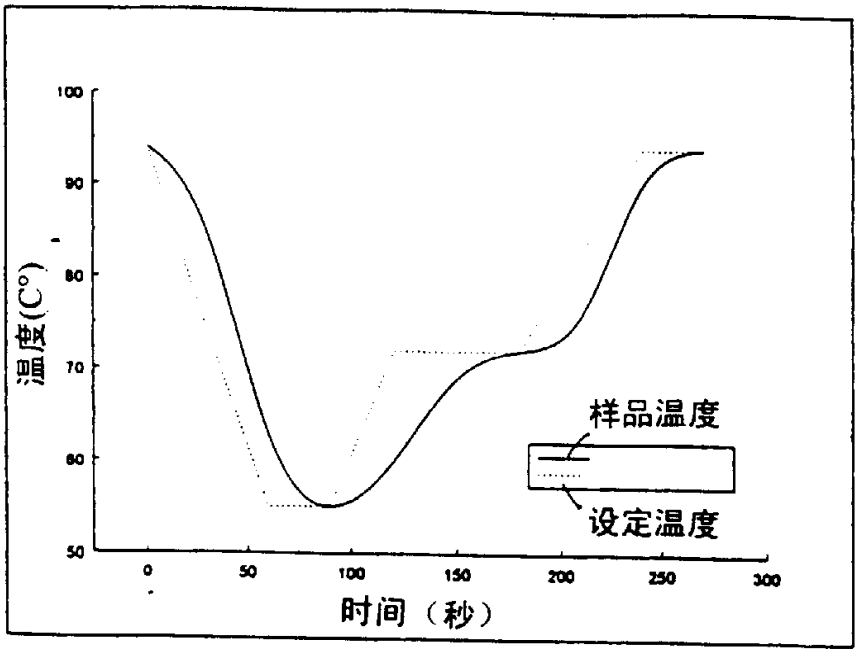


图 3


温度(C°)
100
90
80
70
60
50
0
50
100
150
200
250
300
时间（秒）
样品温度
设定温度

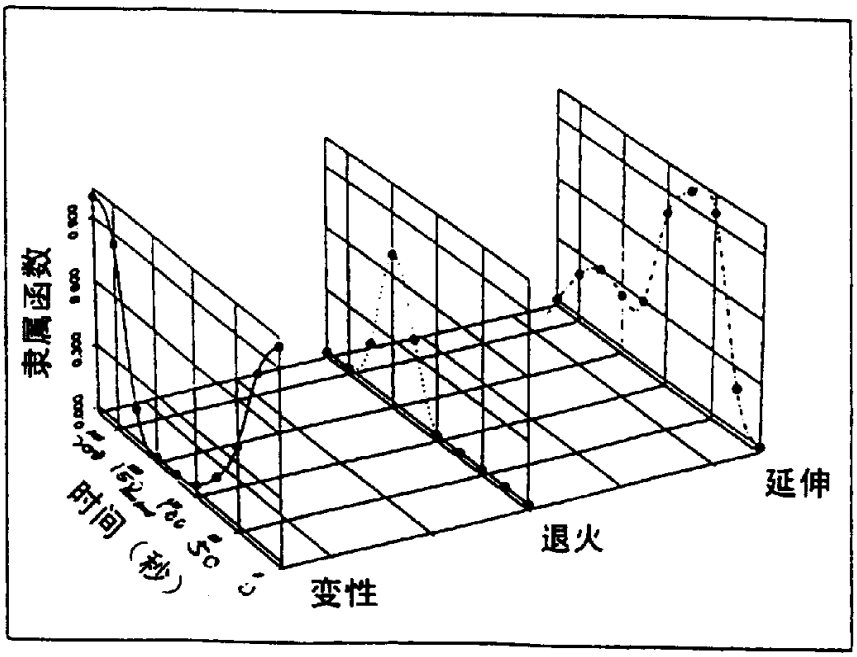

隶属函数
时间（秒）
变性
退火
延伸

图 4

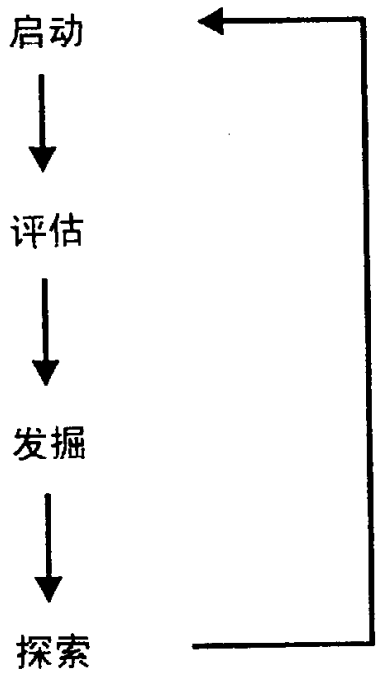

启动
评估
发掘
探索

图 5a

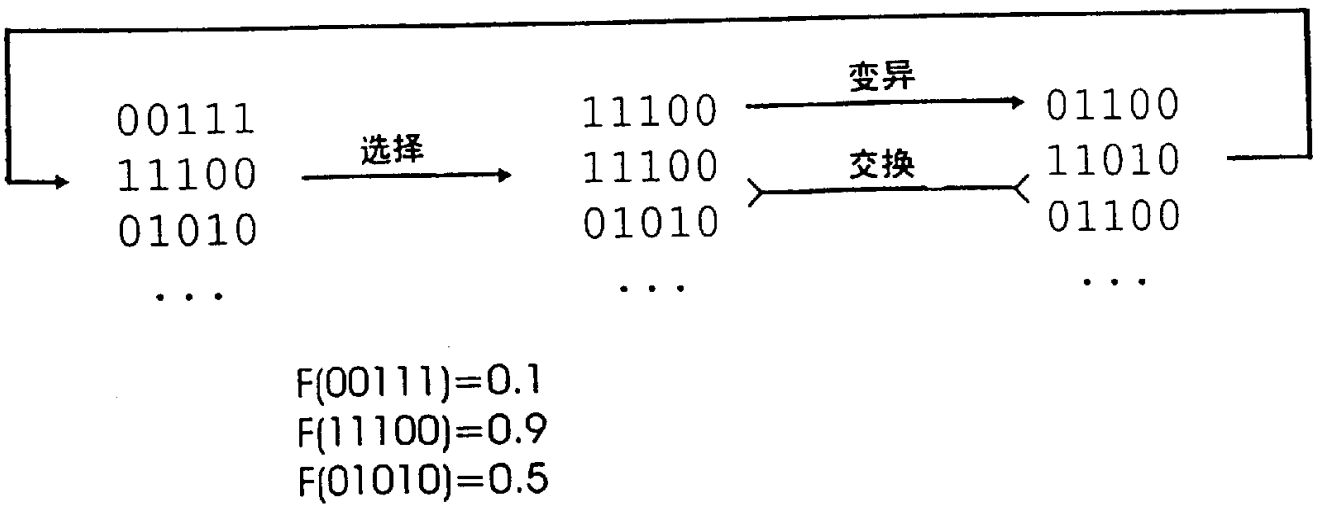

00111
11100
01010
...
选择
11100
11100
01010
...
变异
交换
01100
11010
01100
...
F(00111)=0.1
F(11100)=0.9
F(01010)=0.5

图 5b

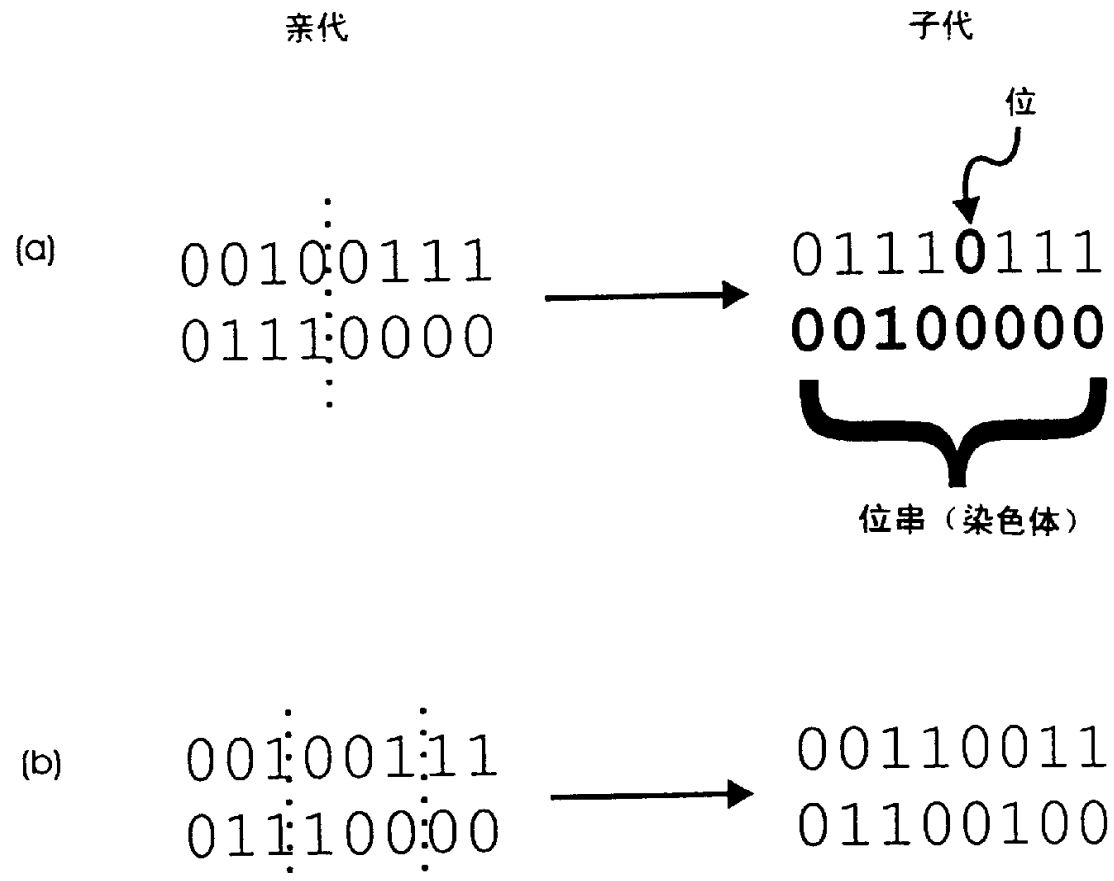
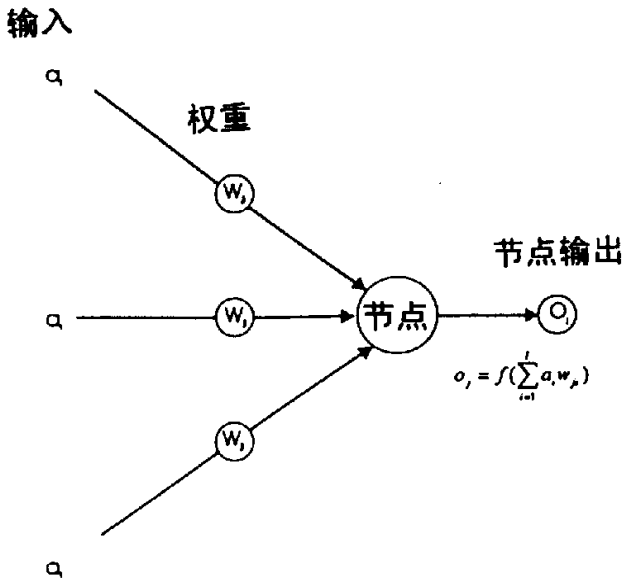

亲代
子代
位
(a)
00100111
01110000
01110111
00100000
位串（染色体）
(b)
00100111
01110000
00110011
01100100

图 6


输入
权重
节点输出
节点
$o_j = f(\sum_{i=1}^{I} a_i w_{ji})$

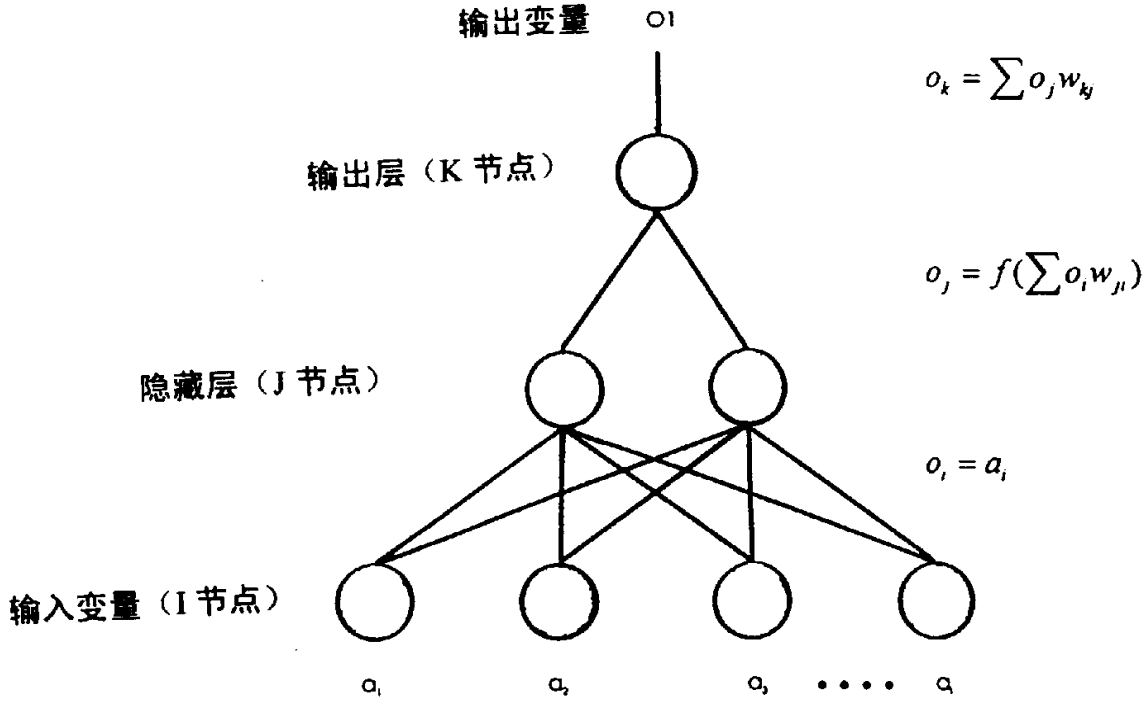

输出变量
O1
$o_k = \sum o_j w_{kj}$
输出层（K 节点）
$o_j = f(\sum o_i w_{ji})$
隐藏层（J 节点）
$o_i = a_i$
输入变量（I 节点）

图 7

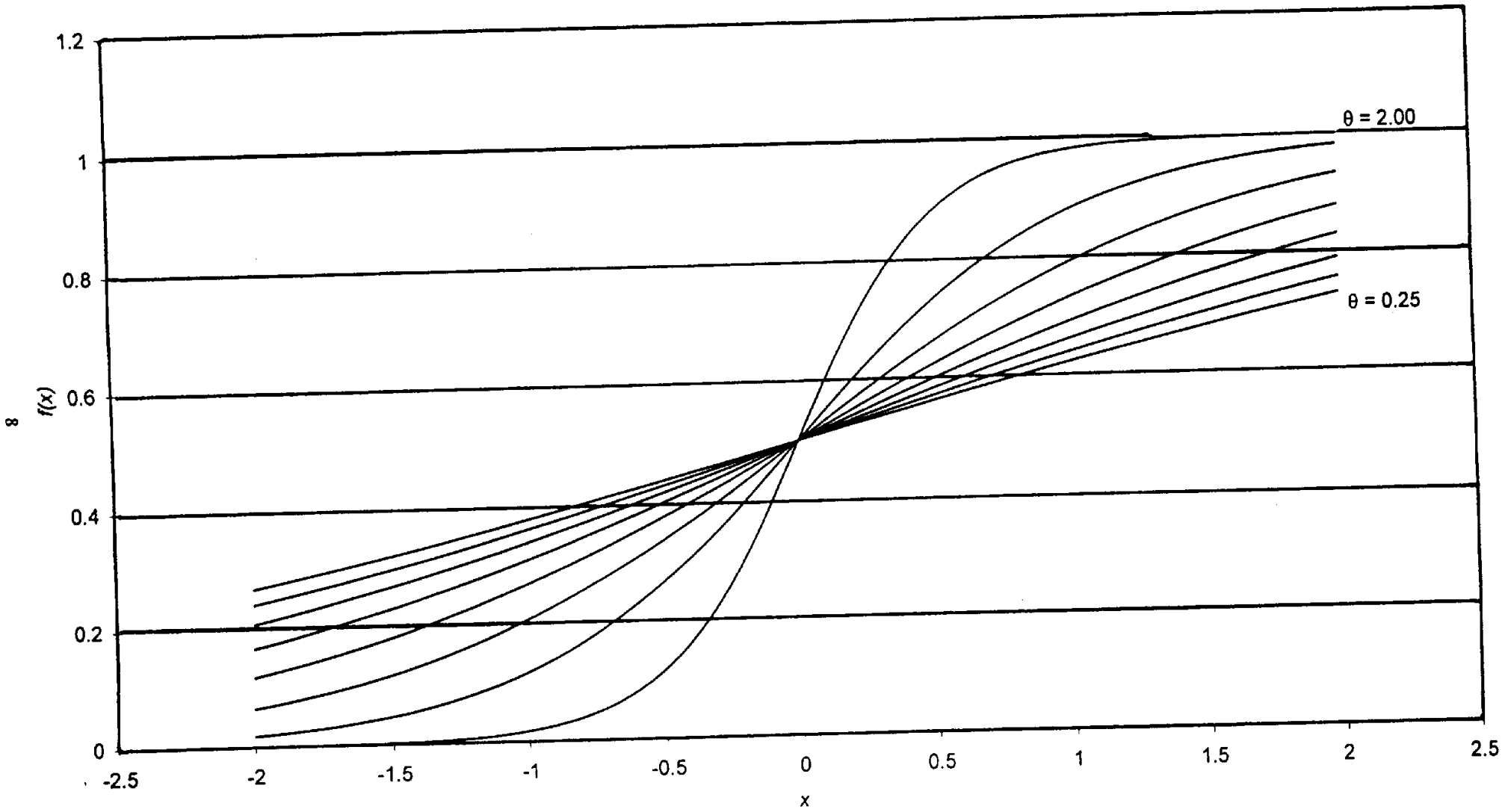


图 8

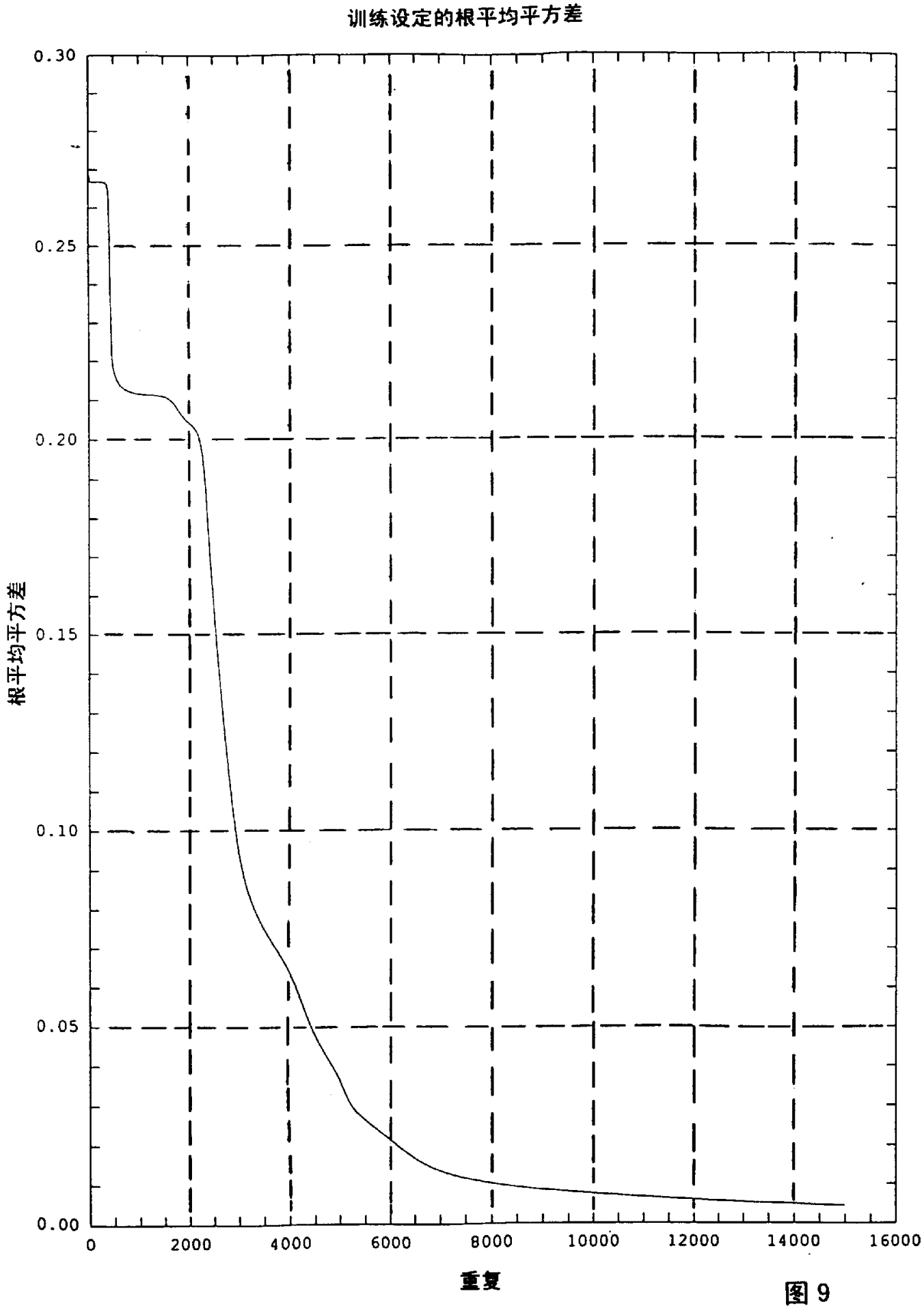

训练设定的根平均平方差
0.30
0.25
0.20
0.15
0.10
0.05
0.00
根平均平方差
0
2000
4000
6000
8000
10000
12000
14000
16000
重复

图 9

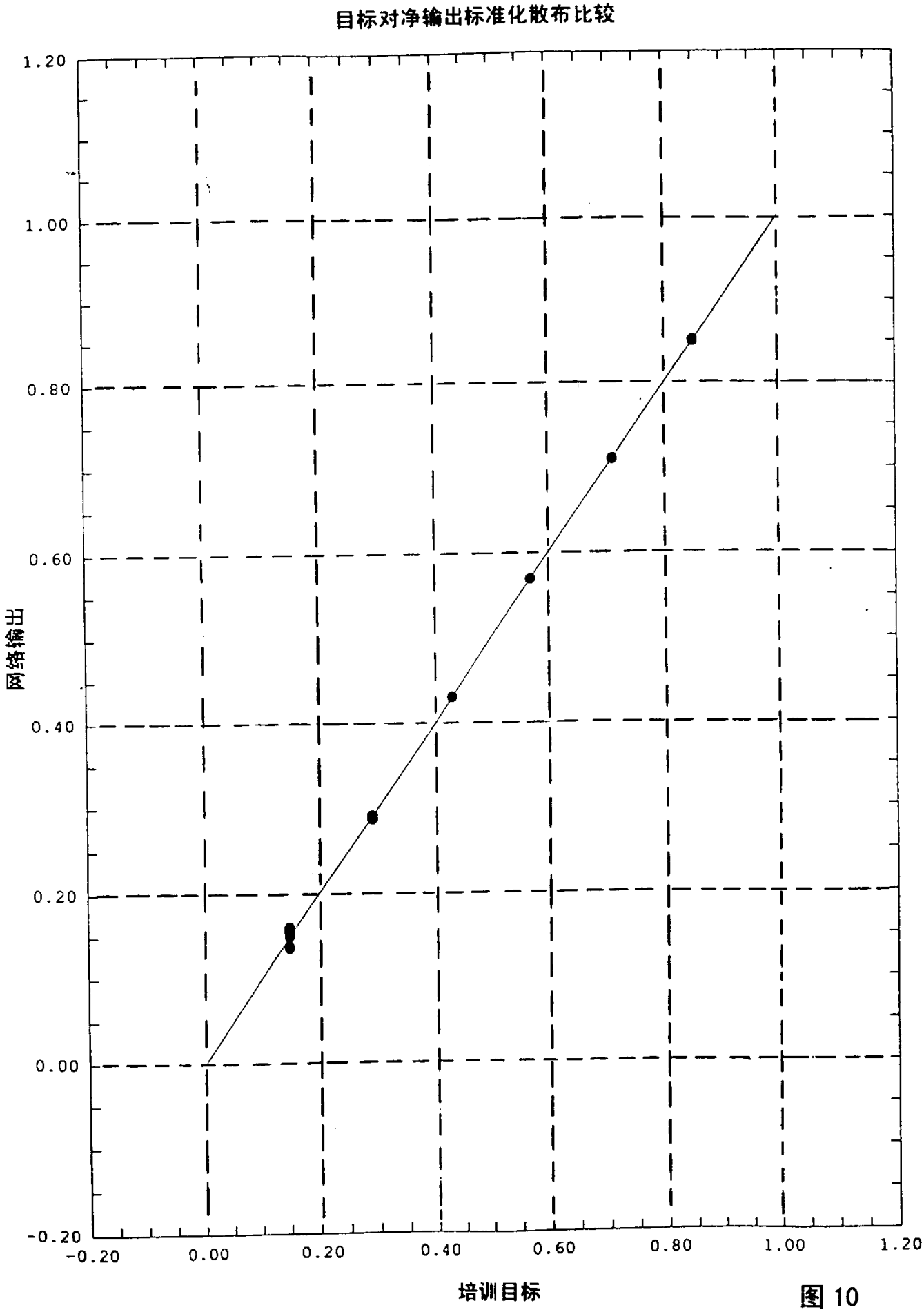

目标对净输出标准化散布比较
网络输出
培训目标
1.20
1.00
0.80
0.60
0.40
0.20
0.00
-0.20
-0.20
0.00
0.20
0.40
0.60
0.80
1.00
1.20

图 10

输入节点 1，2 对输出节点 1 的影响

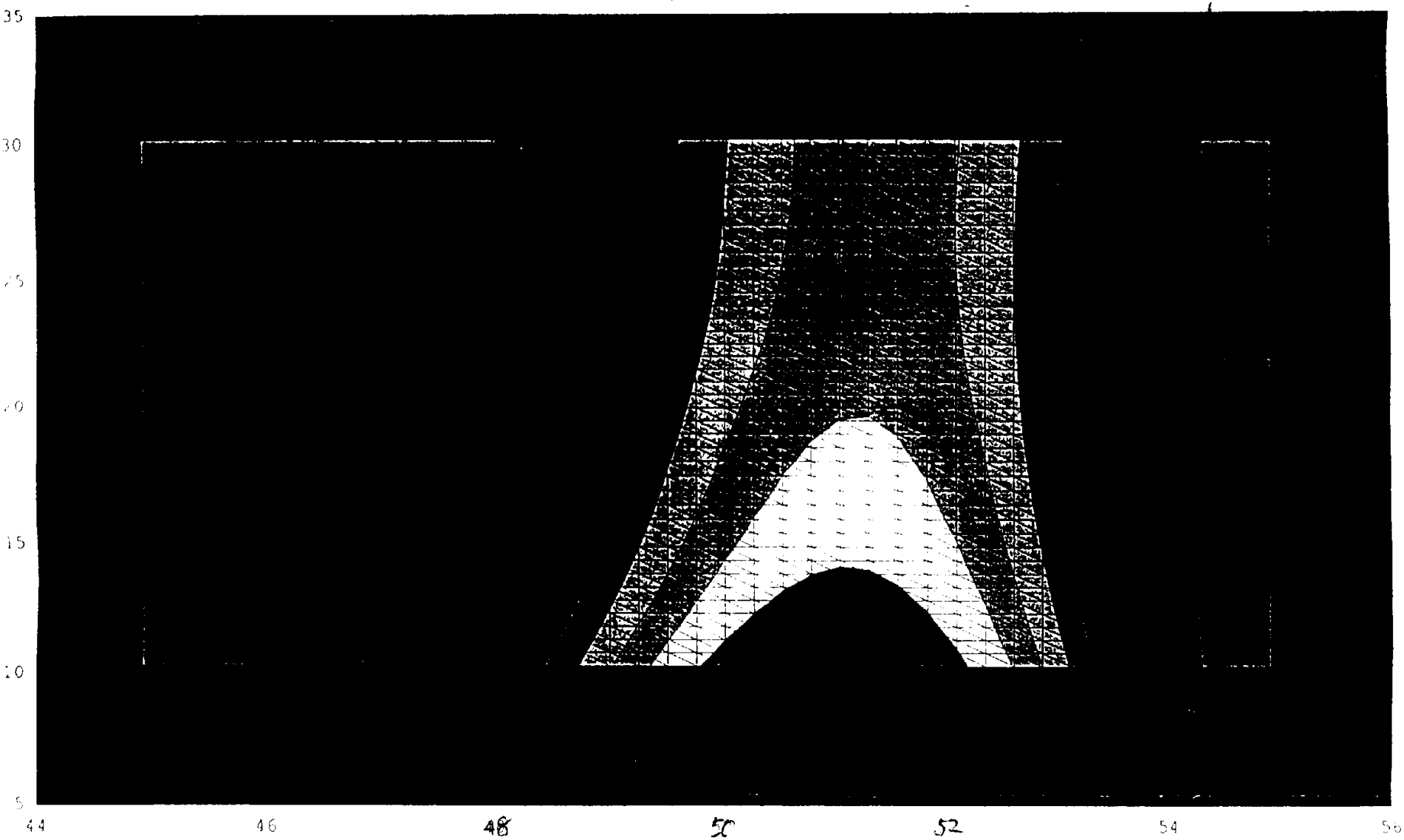


图 二

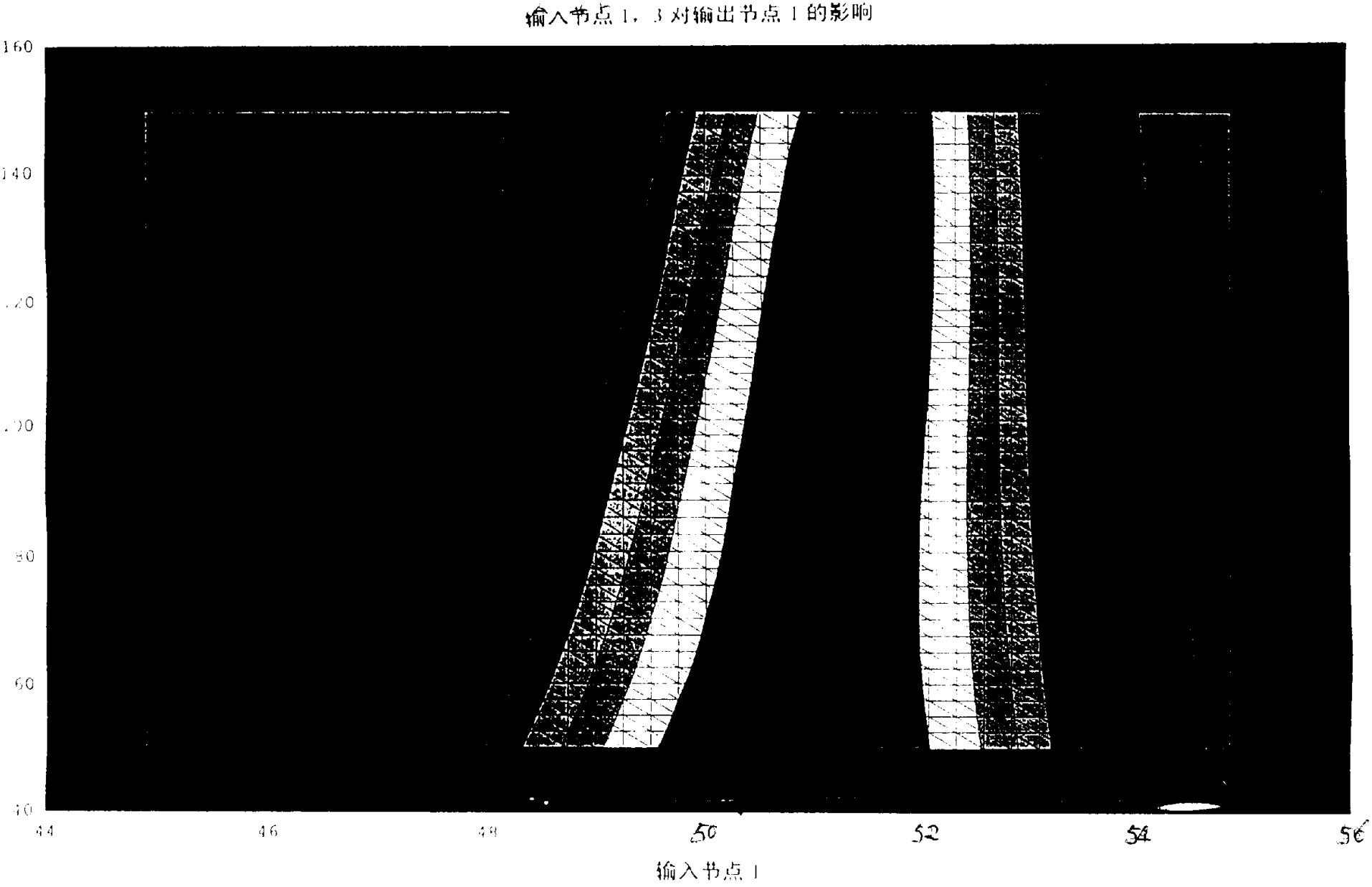

输入节点 1，3 对输出节点 1 的影响
160
140
120
100
80
60
40
44
46
48
50
52
54
56
输入节点 3
输入节点 1

图 12

输入节点 1，4 对输出节点 1 的影响

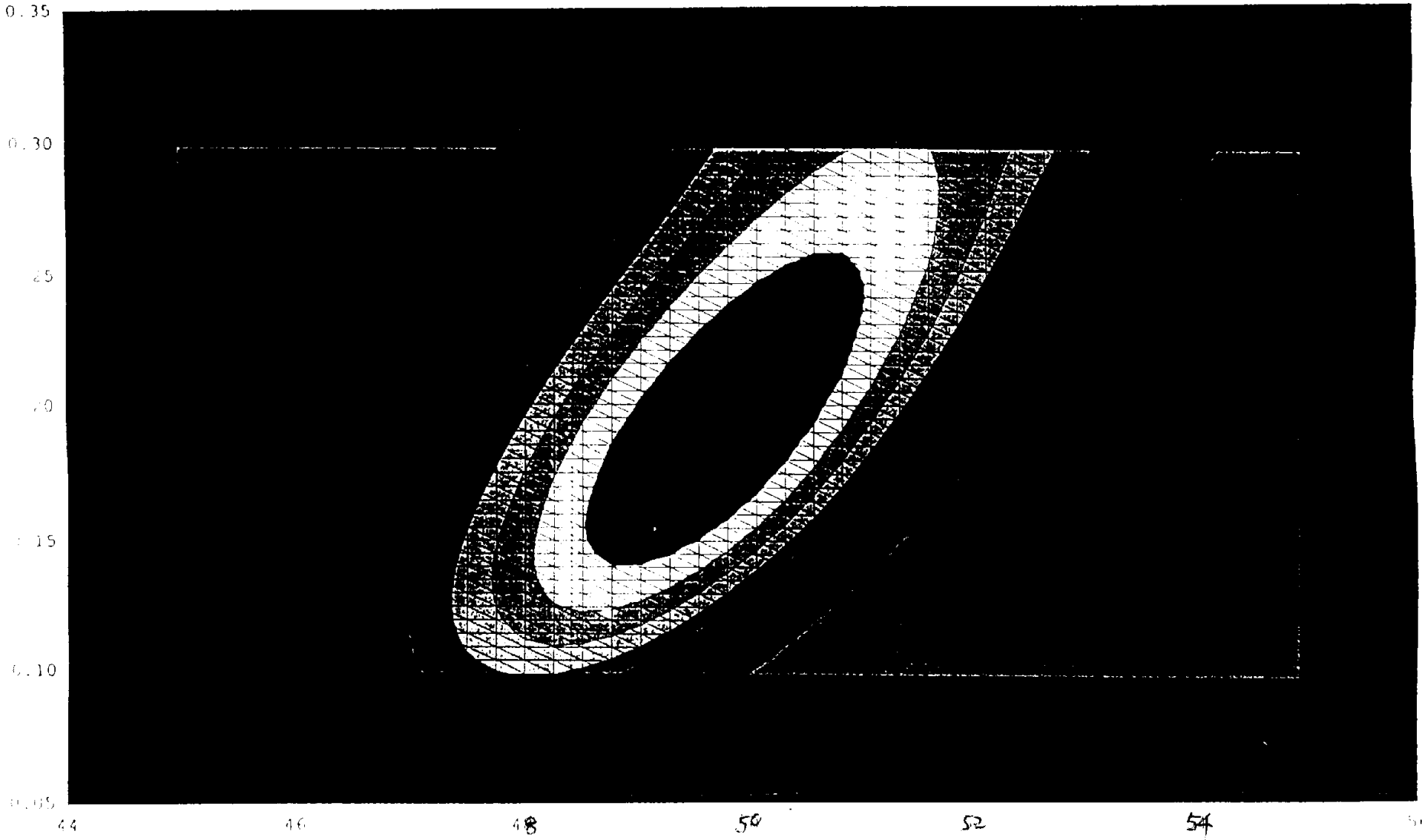


图 13

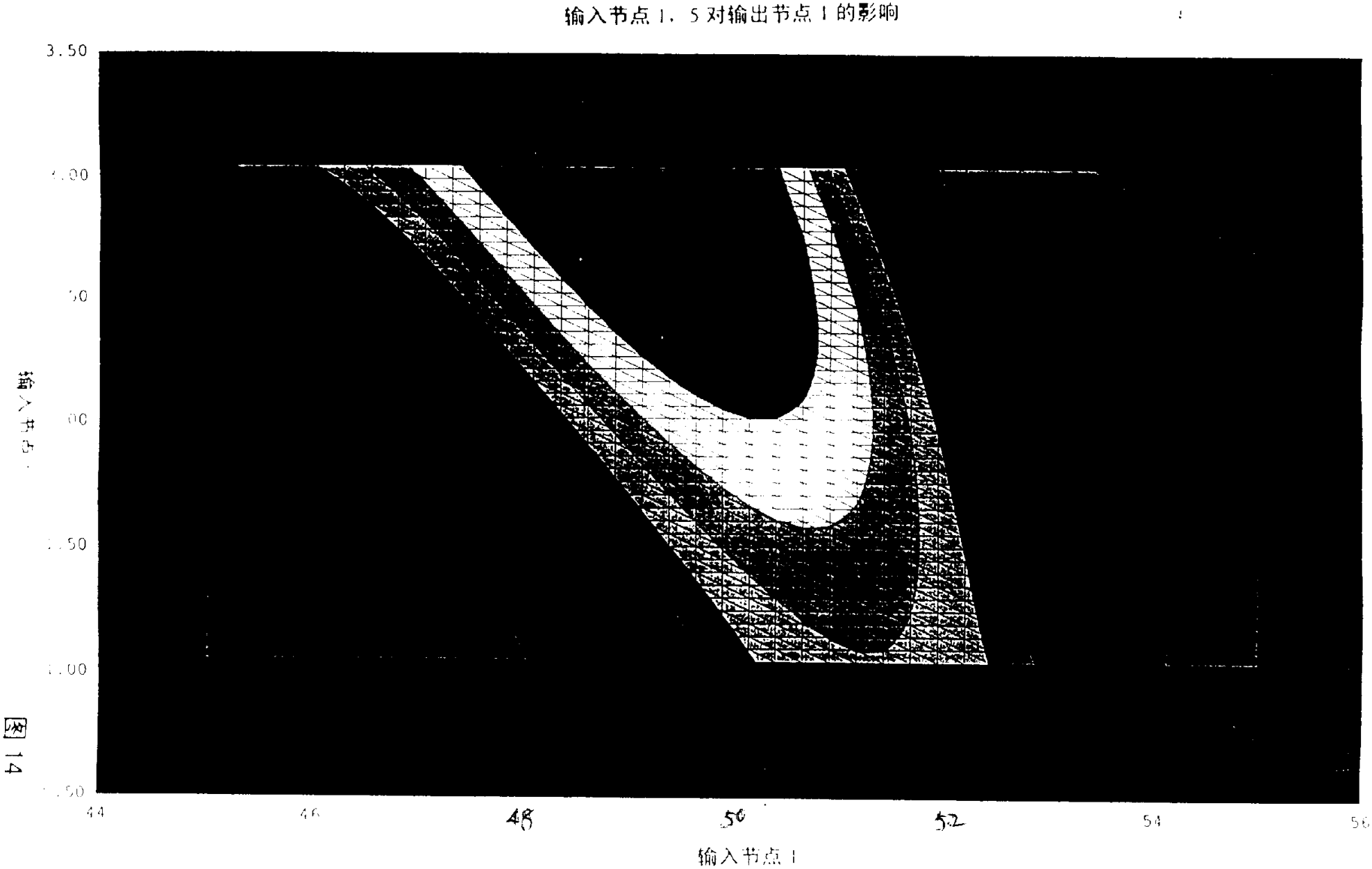

输入节点 1，5 对输出节点 1 的影响
3.50
1.00
.50
.00
1.50
1.00
.50
44
46
48
50
52
54
56
输入节点 1
输入节点 5

图 14

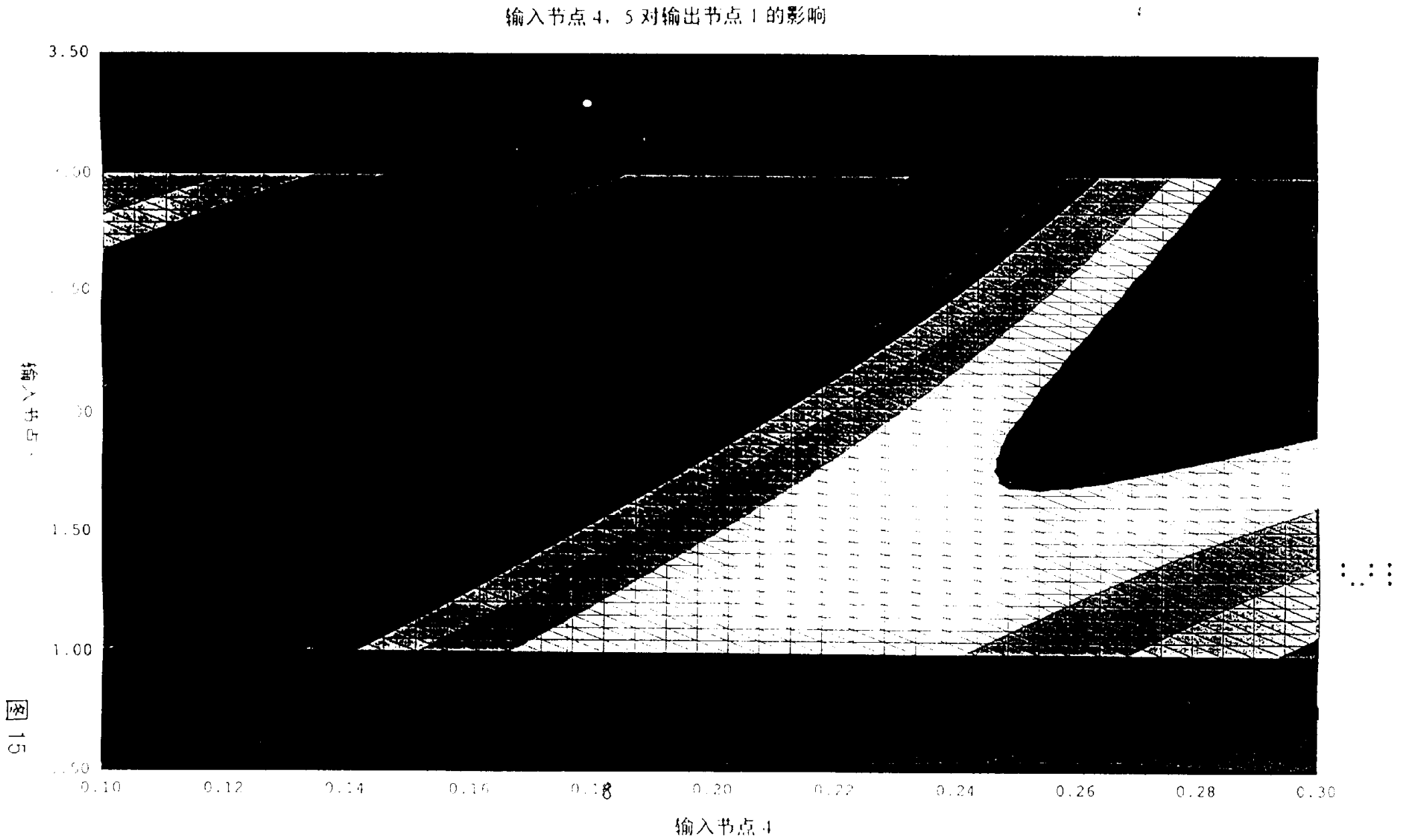

输入节点 4，5 对输出节点 1 的影响
3.50
1.50
1.00
0.10
0.12
0.14
0.16
0.18
0.20
0.22
0.24
0.26
0.28
0.30
输入节点 4

图 15