

Einsatz von Mustererkennungsverfahren zur Optimierung von Chip-Testplänen

Studienarbeit, Universität Karlsruhe
Fakultät für Informatik

Maxim Feinleb

4. Juli 2005

Inhaltsverzeichnis

1	Zusammenfassung	5
2	Problembeschreibung	7
3	Einführung	9
3.1	Stand der Technik	9
3.2	Chip-Tests	10
3.3	Testkostenoptimierung	11
4	Klassifikation	13
4.1	Grundlagen	13
4.2	Klassifikationsverfahren	14
4.2.1	A-Priori-Klassifikatoren	14
4.2.2	Gauß-Klassifikatoren	14
4.2.3	Nächste-Nachbar-Klassifikatoren	16
4.2.4	Multi-Layer Perzeptronen	16
4.3	Kreuzvalidierung	16
4.4	Lineare Diskriminanzanalyse	17
5	Arbeitsumgebung	19
5.1	Experimentdaten	19
5.2	Redundanzanalyse mit reduTec	20
5.2.1	Informationsgehalt	20
5.2.2	Redundanz	22
6	Experimente	23
6.1	Wafer-Klassifikation	24
6.1.1	Klassifikation der Wafer-Ausbeute ohne Verwendung der linearen Diskriminanzanalyse	25
6.1.2	Klassifikation der Wafer-Ausbeute unter Verwendung der linearen Diskriminanzanalyse	27
6.1.3	Precision-Recall Auswertung der Wafer-Ausbeute Klassifikation	28
6.1.4	Fazit: Wafer-Klassifikation	29
6.2	Testplan-Klassifikation	31
6.2.1	Einzeltestvorhersage	32
6.2.2	Testplanauswahl	37
6.3	Kostenrechnung	40

Kapitel 1

Zusammenfassung

Die vorliegende Arbeit verwendet aus der allgemeinen Mustererkennung bekannte Verfahren wie A-Priori- und Bayes-Klassifikation, Gauß-Klassifikatoren, Nächste-Nachbar-Klassifikatoren und Lineare Diskriminanzanalyse zur Klassifikation von Halbleiter-Wafern anhand so genannter PCM-Messwerte, die verschiedene Eigenschaften roher, noch nicht mit den Produkt-ICs belichteten Wafer beschreiben.

Bei korrekter Klassifikation eines Wafers kann vorhergesagt werden, welche Ausbeute an funktionsfähigen Chips zu erwarten ist und ob es sich lohnt diesen Wafer überhaupt weiter zu bearbeiten. Darüber hinaus kann vorhergesagt werden, welche Funktionen der noch auf den Wafer kommenden ICs möglicherweise ausfallen können und daher überprüft werden sollten.

Durch eine geeignete Auswahl von später durchzuführenden Tests - insbesondere durch Weglassen von Tests, die keine Defekterkennung erwarten lassen - können Testzeit und Testkosten eingespart werden. Das durch Weglassen der Tests erhöhte Risiko unentdeckter Defekte und die damit verbundenen Kosten müssen natürlich gegengerechnet werden.

Die vorliegende Arbeit zeigt, dass es möglich ist, durch Analyse von IC-Testplänen und der Klassifikation von Wafern Testzeit einzusparen, sodass die reduzierten Kosten die zusätzlich eingegangenen Risiken deutlich übersteigen.

Anhand historischer Messdaten der Halbleiter-Firma Atmel aus Heilbronn, wurde eine Einsparung von Testkosten in Höhe von ca. 1 bis 5 Euro je Wafer erreicht.

Kapitel 2

Problembeschreibung

Einer der wichtigsten Schritte in der Produktion von ICs ist das Testen der korrekten Funktion derselben. Nach dem heutigen Stand der Technik ist es nicht möglich ICs mit einer ausreichend hohen Wahrscheinlichkeit fehlerfrei zu produzieren. In Abhängigkeit davon, wie lange ein bestimmtes Produkt produziert wird und wie reibungslos die Produktionsprozesse eingefahren sind, beträgt der Ausschuss in der Regel zwischen 5% und 50%. Bei einigen wenigen in den Verkaufszahlen großvolumigen Produkten kann die Ausschussrate auch auf ca. 1% bis 2% gesenkt werden. Aber selbst diese ist viel zu hoch, als dass sie an den Endkunden weitergegeben werden könnte. Daher ist es unvermeidbar, vor der Auslieferung jedes IC zu testen.

IC-Tests verursachen Kosten und benötigen Zeit. Immer wieder kommt es vor, dass das Testen den engsten Flaschenhals in der gesamten Produktion darstellt und so Lieferengpässe ausgelöst werden, oder die Firmen gezwungen sind, für kurze Zeit mehr Tester-Hardware zu betreiben als in Phasen normaler Konjunktur nötig wäre.

So ist es gerade für die europäische Halbleiterindustrie mit ihren im Vergleich zur hauptsächlich asiatischen Konkurrenz relativ hohen Personalkosten wichtig, durch eine höhere Flexibilität und Produktivität erfolgreich zu sein [ITR03]. Die vorliegende Arbeit verfolgt das Ziel, der Halbleiterindustrie ein Mittel zur Reduktion von Testzeiten und Testkosten in die Hand zu geben. Sie wurde daher in enger Kooperation mit der Industrie durchgeführt, insbesondere mit Hilfe der Software der Firma optimiSE in Karlsruhe und Atmel in Heilbronn. Die optimiSE GmbH ist seit mehreren Jahren auf dem Gebiet der Chip-Testplanoptimierung tätig. Die Firma Atmel ist ein weltweiter Halbleiterkonzern mit einer Jahresproduktion von mehreren Milliarden ICs. Durch diese Kooperation und durch die Verwendung realer Daten wird die praktische Relevanz der Arbeit sichergestellt.

Die Produktion und das Testen von Chips besteht aus mehreren Phasen, bestehend aus jeweils mehreren typischerweise Hunderten oder Tausenden einzelner Tests. In der Regel werden die Chips bzw. ihre Ausgangs- und Zwischenprodukte nach jeder Phase getestet, sodass jede Phase ihre eigene Ausschussrate sowie ihre eigenen Produktions- und Testkosten hat. Die meisten

Anstrengungen zur Testzeitreduktion beschränken sich auf einzelne Phasen ohne Berücksichtigung der Testergebnisse vorheriger Phasen. In [Mil98] wird ein Überblick über die heutigen Methoden des Chip-Testens gegeben. Darüber hinaus kann festgestellt werden, dass die meisten Ansätze sich auf die Untersuchung und Interpretation von Fehlermustern konzentrieren, während die in viel größerem Umfang vorhandenen, verschiedenen „guten“ Messwerte nur selten ausgewertet werden. Viele statistische Auswertungen von Messergebnissen betrachten die einzelnen Tests als isolierte Einheiten und gehen meist nicht über die Berechnung und Verfolgung zahlreicher Kenngrößen hinaus. Zu solchen Kenngrößen gehören z.B. Prozessqualitätskriterien wie Mittelwerte, Standardabweichungen, C_{pk} , und so weiter [Cha01].

Die vorliegende Arbeit geht über den Stand der Technik in mehrere Richtungen gleichzeitig hinaus, insofern sie die Tests nicht nur als isolierte Einheiten betrachtet, sondern die Bedeutung eines einzelnen Tests als Bestandteil eines Testplans berücksichtigt und insofern nicht nur der Testplan für eine isolierte Produktionsphase optimiert werden soll, sondern die Messergebnisse der vorhergehenden Phase hinzugezogen werden. Von großer Bedeutung ist dabei die Forderung, dass die resultierenden Verfahren vollautomatisch ablaufen müssen und nicht der Begutachtung oder Überwachung durch Menschen bedürfen. Nur so kann sichergestellt werden, dass tatsächlich beträchtliche zeitliche und finanzielle Einsparungen erreichbar sind. Zur Bewertung der Ergebnisse dieser Arbeit wurden daher die tatsächlichen finanziellen Randbedingungen der Chip-Produktion bei der Firma Atmel zugrunde gelegt.

Kapitel 3

Einführung

Dieses Kapitel gibt einen Überblick über den heutigen Stand der Technik, die Methoden des Chip-Testens und der dabei entstehenden Kosten und Risiken. Danach werden Möglichkeiten der Optimierung von Testplänen nach dem heutigen Stand der Technik beschrieben. Schließlich werden die in dieser Arbeit verwendeten Techniken kurz umrissen.

3.1 Stand der Technik

Die hier vorgestellten Arbeiten betreten ein verhältnismäßig neues Gebiet. Die Problematik der dynamischen Testzeiteinsparung wird in der Praxis bisher nur dadurch gelöst, dass das Testen mit Hilfe eines so genannten *Samplings* durchgeführt wird, bei dem auf jedem Wafer nur ein Teil der ICs komplett getestet wird. Die Auswahl der ICs und der Testpläne erfolgt dabei ohne Berücksichtigung der PCM-Messergebnisse. Die meisten Arbeiten beschäftigen sich mit Chips, die auf analoger Technik basieren oder sie konzentrieren sich auf die Auswertung einfacher statistischer Eigenschaften einzelner Tests (wie C_p und C_{pk}) und reagieren mit der Änderung des Testplans in den Fällen, in denen diese Eigenschaften die vorgesehenen Toleranzbereiche verlassen. Da das gesamte Gebiet der Kostenreduktion im Bereich des Chip-Testens wichtige wirtschaftliche Aspekte tangiert und vor allem dort von Bedeutung ist, wo große Stückzahlen produziert werden, sind die wenigsten Unternehmen bereit Details über die tatsächlichen Wirkungen verschiedener Maßnahmen zu veröffentlichen. Eine weitere Problematik besteht in der Praxis oft darin, dass die einzelnen Phasen der Produktion von Halbleitern in verschiedenen Unternehmen stattfinden, was eine Kommunikation der zur Optimierung der Testpläne benötigten Daten erschwert bzw. oft praktisch unmöglich macht. Umso mehr ist es als Glücksfall zu betrachten, dass für diese Arbeit die Firma Atmel als Kooperationspartner gewonnen werden konnte, die im eigenen Hause an einem Standort verschiedene Entstehungsphasen vereinigt und bereit war, Produktionsdaten zur Verfügung zu stellen.

Da eine Beurteilung der Qualität der Ergebnisse der vorliegenden Arbeit nicht durch Vergleich mit früheren Arbeiten durchgeführt werden konnte,

und da keine allgemein akzeptierten Benchmarks existieren, wurde hier der tatsächlich erzielbare finanzielle Gewinn ermittelt, der durch den Einsatz der erarbeiteten Methoden zu erwarten ist. Die Beurteilung dieser Methoden erfolgt somit durch Vergleich der realen Kosten. Nichtsdestoweniger werden die üblichen Klassifikationsmerkmale und -statistiken wie Erkennungsraten und Precision-/Recall-Werte analysiert.

3.2 Chip-Tests

Der Prozess der Chip-Produktion setzt sich aus mehreren Phasen zusammen, die sich wie in Abb. 3.1 dargestellt, unterteilen lassen. In der ersten Phase werden völlig unbelichtete Wafer erzeugt, dabei fallen Prozess-Messdaten an, wie z.B. gemessene Drücke, Temperaturen, Wafer-Schichtdicken u.s.w. In der zweiten Phase werden vereinzelte (i.d.R. ca. 5 bis 15) Schaltkreise auf den Wafer belichtet und geätzt. Diese Schaltkreise werden später nicht zu ICs, sondern entstehen auf den späteren Sägekanten. Sie dienen lediglich dem Zweck verschiedene elektrische Eigenschaften des Wafers und der auf ihm entstehenden Schaltelemente (z.B. Transistoren, Dioden, Widerstände) messen zu können. Diese kleinen Schaltkreise werden *PCMs* (*Product Control Monitors*) genannt. PCM-Messdaten enthalten mehrere hundert Werte, wie Verstärkungsfaktoren von Transistoren, Größen von Widerständen, Kapazitäten, Schaltverhalten von Dioden etc. Diese geben nicht nur Auskunft über die Gesamtqualität des Wafers sondern auch über die Unterschiede verschiedener Regionen auf dem Wafer.

In der dritten Phase werden die Wafer mit den Produkt-Schaltkreisen (*ICs*) belichtet und geätzt. Bevor diese ausgesägt werden, werden sie einem umfangreichen Test unterzogen, bei dem wiederum hunderte bis über tausend Messwerte erhoben werden. ICs, die die Tests nicht bestehen, werden von der Tester-Hardware automatisch markiert und so vom Rest des Produktionsprozesses ausgeschlossen, damit wird das so genannte *Einpacken* (d.h. Gießen in die *Käferform*) und das danach folgende finale Testen, also die vierte Phase, eingespart. In der fünften Phase werden später komplette Platinen oder gar

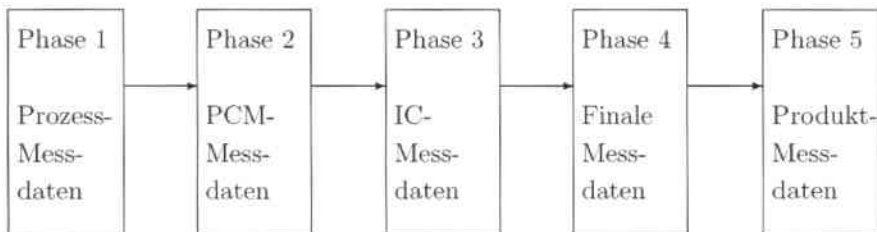


Abbildung 3.1: Chip-Produktion: Prozess in mehreren Phasen

ganze Geräte auf ihre Funktionsfähigkeit getestet.

Testpläne bestehen aus zweierlei Tests, so genannten *parametrischen* und *funktionalen* Tests. Parametrische Tests liefern Messwerte aus einem kontinuierlichen, reellwertigen Merkmalsraum, während funktionale Tests lediglich eine binäre bestanden/nicht-bestanden Messung liefern. Parametrische Tests werden immer zusammen mit einem Intervall angegeben, innerhalb dessen die Messwerte liegen müssen, damit der Test als bestanden gilt.

Eine optimale Teststrategie würde die Auswahl der in jeder Phase durchzuführenden Tests in Abhängigkeit aller bis dahin gemessenen Werte bestimmen. In der Praxis ist dies meist nicht möglich, weil die einzelnen Phasen in verschiedenen Fabriken und verschiedenen Unternehmen stattfinden. Nicht immer ist es möglich von den Wafer-Herstellern alle Prozess- und PCM-Messdaten zu erhalten, noch schwieriger gestaltet sich die Verfolgung einzelner Chip-Identitäten von der Position auf dem Wafer über den Käfer bis hin zur Platine.

Für die vorliegende Arbeit wurden die Phasen 2 und 3 untersucht, da die Firma Atmel in Heilbronn beide durchführt und einander zuordenbare PCM- und IC-Messdaten für diese Arbeit zur Verfügung stellte.

3.3 Testkostenoptimierung

Kosten in der Chip-Herstellung entstehen an vielen Stellen. Die in der vorliegenden Arbeit betrachteten mit dem Testen in Verbindung stehenden Kosten entstehen zum einen durch den Betrieb der Tester-Hardware, die die IC-Messdaten erhebt und zum anderen durch das Verpacken der ausgesägten ICs, die danach noch einmal finalen Tests unterzogen werden.

Typischerweise werden mehrere hundert IC-Tests, die jeweils mehrere Millisekunden dauern pro IC durchgeführt. Ein typischer Wafer wird mit hunderten ICs bestückt, sodass das komplette Testen eines Wafers einige Minuten bis zu über einer Stunde dauern kann. Die Reduktion der Testzeit wirkt sich hier bei einer hohen Auslastung der Produktionsstraßen direkt auf den Durchsatz und die Kosten aus. Durch unnötig durchgeführte Tests entstehen vermeidbare Kosten. Die Identifikation solcher unnötiger Tests erweist sich als durchaus nicht trivial. Zu den fortschrittlichsten Methoden gehört die Analyse des Informationsgehalts der Messwerte eines Tests als Teil eines gesamten Testplans. In [Rog04] wird gezeigt, wie mit der *reduTec* Software der Firma *optimiSE* (siehe Abschnitt 5.2) der relative Informationsgehalt einzelner Tests in Bezug auf die restlichen Tests eines Testplans bestimmt werden kann. Mit Hilfe dieses Informationsmaßes kann ein Ranking der einzelnen Tests berechnet und der allgemeine Testplan in einen obligatorischen und einen erweiterten Testplan aufgeteilt werden. Der erweiterte Testplan würde alle obligatorischen und redundanten Tests beinhalten. Im Normalfall würde stets der obligatorische Testplan eingesetzt werden, falls jedoch kritische oder problematische Bauteile getestet werden müssen, könn-

te der erweiterte Testplan zum Einsatz kommen. Siehe dazu auch Abschnitt 5.2.

Die vorliegende Arbeit verwendet folgende ergänzende Verfahren um die Produktionskosten zu minimieren:

1. Eine Möglichkeit Testkosten zu sparen besteht darin, Wafer mit besonders geringer Ausbeute frühzeitig im Produktionsprozess zu erkennen und aus diesem zu entfernen. Nach der Bestückung der Wafer mit PCMs werden bereits erste Messwerte erhoben. Die Frage, ob die Ausbeute eines bestimmten Wafers unter einem vorgegebenen Schwellenwert liegt und der Wafer somit vom weiteren Produktionsprozess ausgeschlossen werden muss, kann mit Hilfe der gewonnenen Messdaten sowie dem Einsatz eines Klassifikators geklärt werden.
2. Die Testkosten können auch dann eingespart werden, wenn nach dem Abwägen der Risiken festgestellt werden kann, dass ein bestimmter Test auf einem bestimmten Wafer - beispielsweise aufgrund „hervorragender“ PCM-Messergebnisse - nicht ausgeführt werden braucht. Zur Beurteilung der PCM-Messergebnisse wird ebenfalls ein Klassifikator eingesetzt. Die Testauswahl erfolgt dabei auf der Basis des mit Hilfe von reduTec erstellten Rankings der Tests mit geringem Informationsgehalt.

Kapitel 4

Klassifikation

In diesem Kapitel wird zunächst die grundlegende Vorgehensweise beim Klassifizieren bzw. Vorhersagen bestimmter Eigenschaften der Wafer gezeigt. Danach wird auf die verwendeten Klassifikationsverfahren näher eingegangen. Abschließen wird dieses Kapitel mit einer kurzen Vorstellung des Verfahrens der Kreuzvalidierung und der Einführung in lineare Diskriminanzanalyse.

4.1 Grundlagen

Zur Minimierung der Testkosten werden im Rahmen dieser Arbeit mehrere Klassifikationsverfahren eingesetzt. Die Aufgabe aller verwendeter Klassifikatoren ist zum Einen die Aufteilung der Trainingsdaten in zwei Klassen und zum Anderen die Vorhersage einer bestimmten Eigenschaft eines neuen (zu klassifizierenden) Wafers durch die Zuordnung des Wafers zu einer der beiden Klassen.

Die Grundlage für die Bildung der zwei Klassen stellen historische Daten dar. Historische Daten sind Trainingswafer, die sowohl über PCM- als auch IC-Messergebnisse verfügen (siehe Abschnitt 3.2). Werden nun die IC-Messergebnisse einem vorgegebenen Kriterium gegenübergestellt, so können zwei Klassen gebildet werden. In beiden Klassen werden PCM-Messergebnisse zugehöriger Wafer abgelegt. Die Klassen spiegeln danach zwei Stichprobenräume wieder. Der eine Raum beinhaltet Wafer, deren IC-Messergebnisse oberhalb des Kriteriums liegen und der andere Raum beinhaltet die Wafer, deren IC-Messergebnisse unterhalb des Kriteriums zu finden sind.

Das Kriterium steht während der Klassifikation für eine bestimmte zu untersuchende und später vorhergesagte Eigenschaft eines Wafers. Ein solches Kriterium kann beispielsweise die Ausbeute der Wafer sein. Demnach werden zwei Klassen von Trainingswafern erstellt, eine Klasse mit Wafern, deren Ausbeute oberhalb des Schwellenwertes und eine Klasse mit Wafern, deren Ausbeute unterhalb des Schwellenwertes liegt.

Die Vorhersage einer bestimmten Eigenschaft jedes neuen Wafers geschieht durch die Betrachtung der PCM-Messergebnisse des neuen Wafers. Mit unterschiedlichen Verfahren wird auf unterschiedliche Art und Weise untersucht in

welcher der beiden Klassen der vorliegende Wafer mit höherer Wahrscheinlichkeit angehört. Diese Entscheidung wird als Vorhersage der untersuchten Eigenschaft betrachtet.

4.2 Klassifikationsverfahren

Zur Durchführung der eingangs beschriebenen Klassifikation werden mehrere Verfahren eingesetzt. Diese Verfahren sind [DHS00] entnommen und stellenweise auf die vorliegende Problematik optimiert. Die einzelnen Klassifikationsverfahren werden im Folgenden genau beschrieben, in den Abschnitten 6.1 und 6.2 miteinander verglichen und zusätzlich dem A-Priori-Klassifikator gegenübergestellt.

4.2.1 A-Priori-Klassifikatoren

Der A-Priori-Klassifikator baut auf einer sehr einfachen Logik auf. Dieser Klassifikator benötigt keine weiteren Informationen, außer der Stichprobengröße beider Klassen. Aus der Stichprobengröße einer Klasse berechnet er die relative Wahrscheinlichkeit dieser Klasse:

$$P(\text{Klasse}_{(\text{Kriterium} \geq \text{Schwellenwert})}), P(\text{Klasse}_{(\text{Kriterium} < \text{Schwellenwert})})$$

Die Vorhersage über die Zugehörigkeit zu einer Klasse wird einzig auf Basis der folgenden Bedingung getroffen:

$$P(\text{Klasse}_{(\text{Kriterium} \geq \text{Schwellenwert})}) \geq P(\text{Klasse}_{(\text{Kriterium} < \text{Schwellenwert})})$$

Gilt diese, so wird als Vorhersage stets die $\text{Klasse}_{(\text{Kriterium} \geq \text{Schwellenwert})}$ gewählt. Gilt diese Bedingung nicht, so wählt der Klassifikator immer die $\text{Klasse}_{(\text{Kriterium} < \text{Schwellenwert})}$.

Wählt man also beispielsweise die Trainingsdaten so, dass es mehr Stichproben gibt, die das Kriterium erfüllen (oberhalb dieses liegen), wählt der Klassifikator bei einem neuen (zu klassifizierenden) Wafer die $\text{Klasse}_{(\text{Kriterium} \geq \text{Schwellenwert})}$ ohne Beachtung der tatsächlichen PCM-Messwerte. Aus genau diesen Grund werden alle folgenden Klassifikationsverfahren C_i mit diesem „naiven“ Verfahren in den Abschnitten 6.1 und 6.2 verglichen um einen möglichen Informationsgewinn festzustellen:

$$\Delta I(C_i) = |\text{Fehlerrate}(A\text{-Priori-Klassifikator}) - \text{Fehlerrate}(C_i)| \quad (4.1)$$

4.2.2 Gauß-Klassifikatoren

Die Gauß-Klassifikatoren lassen sich ausschließlich dann einsetzen, wenn der gesamten Stichprobe eine Normalverteilung unterstellt wird. In diesem Fall kann

bei einfacher, eindimensionaler Verteilung die Wahrscheinlichkeit für die Zugehörigkeit eines neuen Wafer x zu der Klasse i durch die bekannte Funktion der Wahrscheinlichkeitsverteilung bestimmt werden:

$$p_i(x) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x - \mu_i}{\sigma_i} \right)^2 \right] \quad (4.2)$$

Der Erwartungswert μ_i und die Standardabweichung σ_i sind natürlich abhängig von der gewählten Klasse.

Zur Erhöhung numerischer Stabilität lässt sich die Funktion 4.2 wie folgt abwandeln:

$$p_{i_{\ln}}(x) = -\ln \left(\sigma_i \sqrt{2\pi} \right) - \frac{1}{2} \left(\frac{x - \mu_i}{\sigma_i} \right)^2$$

Wie bereits gesehen, haben die betrachteten PCM-Messergebnisse eines Wafers jedoch bis zu einigen hundert Werte, wodurch die Verwendung des multivariaten Gauß-Klassifikators nötig wird. Die Wahrscheinlichkeitsfunktion hat dann für die Klasse i und den PCM-Messergebnis-Vektor \vec{x} eines neuen Wafers die Form:

$$p_i(\vec{x}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_i|}} \exp \left[-\frac{1}{2} (\vec{x} - \vec{\mu}_i)^t \Sigma_i^{-1} (\vec{x} - \vec{\mu}_i) \right] \quad (4.3)$$

wobei d der Dimension des Messergebnis-Vektors \vec{x} , $\vec{\mu}_i$ dem d -dimensionalen Mittelwert-Vektor und Σ_i der $(d \times d)$ Kovarianz-Matrix der Klasse i entspricht.

Zur Erhöhung der numerischen Genauigkeit der Funktion 4.3 lässt sich diese ebenfalls wie folgt anpassen:

$$p_{i_{\ln}}(\vec{x}) = -\frac{1}{2} \ln \left[(2\pi)^d |\Sigma_i| \right] - \frac{1}{2} \left[(\vec{x} - \vec{\mu}_i)^t \Sigma_i^{-1} (\vec{x} - \vec{\mu}_i) \right]$$

Die letztlich für die Vorhersage ausschlaggebende Bedingung:

$$P_{Klasse_1}(\vec{x}) \geq P_{Klasse_2}(\vec{x}) \quad (4.4)$$

bzw.

$$P_{\ln_{Klasse_1}}(\vec{x}) \geq P_{\ln_{Klasse_2}}(\vec{x})$$

vergleicht die Wahrscheinlichkeiten für die Zugehörigkeit des PCM-Messergebnis-Vektors \vec{x} zu einer der beiden Klassen und entscheidet sich für die Klasse mit der höheren Wahrscheinlichkeit.

Die Bedingung 4.4 kann zusätzlich um einen klassenspezifischen Gewichtungsfaktor w_i ergänzt werden:

$$w_1 P_{Klasse_1}(\vec{x}) \geq w_2 P_{Klasse_2}(\vec{x}) \quad (4.5)$$

Die Anzahl potentiell unentdeckter Defekte lässt sich möglicherweise verringern, indem die Klasse ($Kriterium < Schwellenwert$) eine stärkere Gewichtung bekommt. Gleichzeitig bedeutet das aber auch eine höhere Wahrscheinlichkeit für die Entstehung von *false positives*.

4.2.3 Nächste-Nachbar-Klassifikatoren

Die Nächste-Nachbar-Klassifikation kann eingesetzt werden, ohne dass bestimmte Annahmen über die der Stichprobe zugrunde liegende Verteilung gemacht werden. Stattdessen lässt sich mit diesem Klassifikationsverfahren bei der Betrachtung der PCM-Messergebnisse eines neuen Wafers die Entscheidung über die Zugehörigkeit des Wafers zu einer Klasse augenblicklich treffen.

Zur Beurteilung der Wahrscheinlichkeit kann als Metrik die euklidische Distanz verwendet werden:

$$e(\vec{x}, \vec{y}) = \|\vec{x} - \vec{y}\| = \sqrt{(x_1 - y_1)^2 + \dots + (x_d - y_d)^2} \quad (4.6)$$

Zur Optimierung der Rechenzeit, lässt sich statt dieser euklidischen Distanz ebenfalls ihr Quadrat verwenden.

Zu einem gegebenen PCM-Messergebnis-Vektor \vec{x} werden nun mit Hilfe einer Distanz-Metrik (bspw. 4.6) die k nächstgelegenen PCM-Messergebnis-Vektoren \vec{p}_α , $\alpha \in \{1, \dots, k\}$ ermittelt. Anschließend werden die Wahrscheinlichkeiten für die Zugehörigkeit zu einer Klasse K_i berechnet:

$$p_i(\vec{x}) = \frac{1}{k} \sum_{\substack{n_i \leq k \\ p_\theta \text{ in } K_i}} \vec{p}_\theta \quad (\vec{p}_\theta \in \vec{p}_\alpha)$$

Dabei entspricht n_i der Anzahl der nächstgelegenen PCM-Messergebnis-Vektoren \vec{p}_θ aus der i . Klasse.

Wie auch bei den Gauss-Klassifikatoren sind für die Nächste-Nachbar-Klassifikatoren die Bedingungen 4.4 und 4.5 für die Vorhersage entscheidend.

4.2.4 Multi-Layer Perzeptronen

In dieser Arbeit wurde zusätzlich ein künstliches neuronales Netzwerk zur Vorhersage einer bestimmten Eigenschaft eines Wafers eingesetzt. Hierbei wurden mehrlagige Perzeptronen verwendet mit variierender Anzahl versteckter Schichten (*hidden layers*) und variierender Anzahl von Neuronen. Die Trainingsdaten wurden mit den neuronalen Netzen mehrere Epochen geübt und schließlich zur Vorhersage von neuen Wafers verwendet.

Als Implementierung wurde eine fertige OpenSource Bibliothek *NeuroBox* [Rue04] eingesetzt.

4.3 Kreuzvalidierung

Für die in Abschnitten 6.1 und 6.2 beschriebenen Experimente wird die *Kreuzvalidierung* verwendet. Die Kreuzvalidierung ist ein statistisches Verfahren, mit welchem die zur Verfügung stehende Datenmenge, bestehend aus n Stichproben, in n Teilmengen aufgeteilt wird. Dabei werden $n - 1$ Proben als Trainingsdaten verwendet und eine Probe zum Klassifizieren benutzt. Es werden anschließend n

Testdurchläufe durchgeführt. Die Gesamtfehlerquote errechnet sich als Durchschnitt aus den Einzelfehlerquoten der n Testdurchläufe. Das beschriebene Verfahren wird separat auf beiden Testsätzen (siehe Abschnitt 5.1) von Wafern durchgeführt.

4.4 Lineare Diskriminanzanalyse

Die lineare Diskriminanzanalyse ist ein bekanntes Verfahren das häufig eingesetzt wird um die Anzahl der Dimensionen des Stichprobenraumes zu reduzieren (siehe dazu auch [Fuk90] und [DHS00]). Der bei diesem Verfahren verfolgte Ansatz ist die Abbildung eines hochdimensionalen Raumes auf einen niederdimensionalen Raum. Bei einem Zwei-Klassen-Problem ist der Wertebereich der Abbildung eindimensional, also eine Gerade.

Die Problematik besteht darin, die Abbildung so zu gestalten bzw. die Gerade derart zu legen, dass die Punktwolken beider Klassen möglichst weit voneinander und mit möglichst geringer Überschneidung auf der Geraden erscheinen.

Angenommen es existiert ein Satz von n Trainingsdaten und somit auch von n d -dimensionalen PCM-Messergebnissen bzw. Stichproben x_1, \dots, x_n . Dieser Satz lässt sich in zwei Klassen aufteilen K_1 und K_2 . Die gewünschte Abbildung soll die folgende Form haben:

$$y = w^t x \quad (4.7)$$

Nach der von Fisher durchgeführten linearen Diskriminanzanalyse ergibt sich der Transformationsvektor w als:

$$w = S_W^{-1}(m_1 - m_2)$$

Wobei m_i den d -dimensionalen Mittelwert der i . Klasse und S_W die Summe der Streuungsmatrizen S_i beider Klassen darstellt:

$$\begin{aligned} S_W &= S_1 + S_2 \\ S_i &= \sum_{x \in K_i} (x - m_i)(x - m_i)^t \end{aligned}$$

Setzt man den erhaltenen Transformationsvektor w in 4.7 ein, so erhält man eine lineare Abbildung mit maximalem Verhältnis der klassenexternen Streuung (zwischen den beiden Klassen) zur klasseninternen Streuung (innerhalb der Klassen).

Zu den Vorteilen der linearen Diskriminanzanalyse gehört einerseits die Verringerung der Komplexität der ihr folgenden Klassifikation und andererseits - und vor allem für bessere Klassifikationsergebnisse wichtig - die Eliminierung unnötiger, keine Information enthaltender Dimensionen. Ein direkter Vergleich der Klassifikationsergebnisse mit und ohne einer vorher durchgeführten linearen

Diskriminanzanalyse ergibt (auf den im Abschnitt 5.1 vorgestellten Testdaten) eine Verbesserung der Erkennungsrate um bis zu 14 Prozent.

Kapitel 5

Arbeitsumgebung

In diesem Kapitel werden sowohl die für die Experimente verwendeten Datensätze als auch generell die Programmumgebung `reduTec` näher betrachtet.

5.1 Experimentdaten

Die in den Experimenten (siehe Abschnitte 6.1 und 6.2) verwendeten Daten lassen sich in zwei Datensätze aufspalten. Die Daten stammen von der bereits erwähnten in Heilbronn ansässigen Halbleiter-Firma Atmel. Die Daten enthalten neben den PCM-Messdaten auch die zugehörigen IC-Messergebnisse. Die Messergebnisse wurden in den im Abschnitt 3.2 vorgestellten Phasen den Wafern entnommen.

An den oben erwähnten Experimenten nehmen insgesamt 271 Wafer von 12 Losen¹ teil. Der erste Testsatz T_1 enthält 103 Wafer und der zweite Testsatz T_2 168 Wafer. Jeder Testsatz repräsentiert hierbei ein bestimmtes Produkt der Firma Atmel.

Jeder Wafer enthält 5 PCMs, die Anzahl der ICs schwankt jedoch, je nachdem welcher Testsatz betrachtet wird. Im T_1 enthält ein Wafer im Durchschnitt 3200 ICs und im T_2 2600 ICs. Jeder PCM beider Testsätze hat im Schnitt 89 PCM-Messergebnisse. Die Anzahl der IC-Messergebnisse in T_1 liegt bei 87, in T_2 bei 129. Daraus folgt, dass in dieser Arbeit 1355 PCMs und 769 000 ICs betrachtet werden. Die Summe aller analysierten PCM-Messergebnisse beträgt mehr als 120 000 und die Summe der IC-Messergebnisse mehr als 85,3 Mio.

Für die Berechnung des Kostenaufwandes für das Testen beider Testsätze in der 3. Phase des Produktionsprozesses werden die durchschnittlichen Kosten der Firma Atmel angenommen. Demnach kostet das Testen eines ICs mit einem Test circa 0,0001 EUR. Damit ergibt sich bei 85,3 Mio. Tests ein Betrag von über 8500 EUR für 271 Wafer. Das Testen eines Wafers aus T_1 kostet demnach knapp 28 EUR, eines Wafers aus T_2 mehr als 33 EUR.

¹Ein Los bezeichnet dabei ein zusammenhängendes Stück Silizium welches im Normalfall in 25 Wafer zerschnitten wird.

5.2 Redundanzanalyse mit reduTec

Von der Firma optimiSE wurde das Software-Tool reduTec [RvSK05], [Rog04] für die Redundanzanalyse von Testplänen zur Verfügung gestellt. Die Identifikation von redundanten Tests als Kandidaten für das Weglassen aus einem langen Testplan zur Generierung eines kurzen Testplans ist die Voraussetzung für die weiteren Experimente.

5.2.1 Informationsgehalt

Im Allgemeinen bedeutet Redundanz in einem Testplan, dass ein Test t als möglicherweise weglassbar betrachtet werden kann, da eine Menge $S(t) = t_1, t_2, \dots$ von Tests existiert, sodass immer dann, wenn t einen Messwert außerhalb des Toleranzbereichs liefert, mindestens ein Test aus $S(t)$ auch einen Messwert außerhalb des Toleranzbereichs liefert. So könnte ein defektes IC niemals lediglich durch Nichtbestehen des Tests t erkannt werden, sondern hätte immer mindestens einen weiteren nicht bestandenen Test. Gerade bei Produkten, die mit sehr hoher Ausbeute produziert werden, ist die Menge an defekten Bauteilen so gering, dass eine bloße Betrachtung der Tests auf defekten Bauteilen zu statistisch insignifikanten oder gar irreführenden Ergebnissen führen würde. Daher konzentriert sich die reduTec Analyse vor allem auf die Messwerte der bestandenen Tests.

Bezeichne $t(d)$ den Messwert von Test t auf dem IC d . Man kann nun sagen, dass t keine Information liefert, wenn der Vektor $t(d)$ aller Messwerte von t eine Linearkombination anderer Vektoren $t_1(d), t_2(d), \dots$ ist. In der Praxis kommt eine derartige exakte Gleichheit einer Linearkombination für parametrische Tests so gut wie nie vor. Es müssen daher weitere Kriterien zur Bestimmung der Redundanz betrachtet werden.

In reduTec wird der normalisierte Messwert $t'(d)$ für einen Test mit den tolerierten Intervallgrenzen $L(t)$ und $H(t)$ und dem Bestwert $\mu(t)$ definiert als:

$$t'(d) = \begin{cases} 1 & \text{für } t(d) < L(t) \\ \left(\frac{\mu(t) - t(d)}{\mu(t) - L(t)} \right)^2 & \text{für } L(t) \leq t(d) \leq \mu(t) \\ \left(\frac{\mu(t) - t(d)}{\mu(t) - H(t)} \right)^2 & \text{für } \mu(t) \leq t(d) \leq H(t) \\ 1 & \text{für } t(d) > H(t) \end{cases} \quad (5.1)$$

Diese Normalisierung wird in Abb. 5.1 dargestellt. Offensichtlich ist $t'(d) = 0$, wenn der gemessene Wert $t(d)$ gleich dem Bestwert $\mu(t)$ ist und $t'(d) = 1$ für alle Messwerte außerhalb des Toleranzintervalls gilt. Durch die Quadrierung werden Abweichungen vom Bestwert und Annäherungen an die Intervallgrenzen besonders herausgehoben.

Man kann nun einen Test t_1 als „kritischer“ auf dem Bauteil d als Test t_2 definieren, wenn $t_1'(d) > t_2'(d)$.

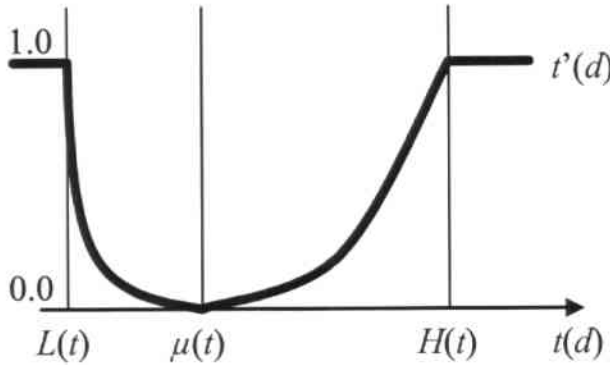


Abbildung 5.1: Ergebnis-Normalisierung

Wenn

$$\forall y \forall d \in D : t_x t(d) \leq t_y t(d), \quad (5.2)$$

dann bedeutet dies, dass t_x keinen praktischen Nutzen zur Fehlererkennung auf der Bauteilemenge D hat, weil für jeden einzelnen Messwert ein anderer Test kritischer ist. Diese ist insbesondere für jeden Test der Fall, der nur Bestwerte misst. Demnach ist ein Test, der immer nur Beinahe-Bestwerte misst weniger kritisch als einer, dessen Werte weiter weg vom Bestwert liegen.

reduTec weist so einem Test der immer nur Bestwerte misst (d.h. $\forall d t(d) = 0$) den Informationsgehalt 0 zu. Ein Test t_1 , der nur für Bauteil d den Nicht-Bestwert $t_1 t(d)$ misst, hat den zusätzlichen Informationsgehalt r_1 ($0 < r_r \leq 1$) gegenüber t_0 . Ein anderer Test t_2 , der sich wie t_1 verhält bis auf $t_2 t(d) = r_2$ ($r_1 \leq r_2 \leq 1$), hat dann einen zusätzlichen Informationsgehalt von $r_2 - r_1$ gegenüber t_1 . Im Allgemeinen gilt: Der Informationsgewinn auf einer Bauteilemenge D durch einen Test t_x , der die Werte $t_x(1), t_x(2), \dots, t_x(|D|)$ misst, gegenüber einem Test t_y , der die Werte $t_y(1), t_y(2), \dots, t_y(|D|)$ misst, ist definiert als:

$$\Delta I(t_x, t_y) = \sum_{d \in D} \min(t_x t(d) - t_y t(d), 0) \quad (5.3)$$

Der Informationsgewinn eines Tests t_x gegenüber einer Menge von Tests $T = t_1, t_2, \dots, t_K$ mit den Messwerten $t_1(1), t_1(2), \dots, t_1(|D|), t_2(1), \dots, t_2(|D|), \dots, t_K(1), \dots, t_K(|D|)$ ist entsprechend:

$$\Delta I(t_x, t_1, t_2, \dots, t_K) = \sum_{d \in D} \min(t_x t(d) - \max_{k=1}^K t_k t(d), 0) \quad (5.4)$$

Das Ranking der Tests geschieht durch den folgendem Algorithmus:

1. Wähle als Nummer eins der Liste einen Test mit dem kleinsten individuellen Informationsgehalt.

2. Wenn k Tests t_1, t_2, \dots, t_k bereits einen Rang haben, dann wähle den nächsten Test t_z , sodass

$$z = \operatorname{argmin}_{x \in \{1 \dots k\}} \Delta I(t_x, \{t_1, t_2, \dots, t_k\}) \quad (5.5)$$

3. Wiederhole Schritt 2 bis alle Tests einen Rang haben.

Dieser Ansatz stellt sicher, dass der Test mit dem Rang k den geringsten Informationsabstand gegenüber der Menge der Tests mit dem Rang $< k$ besitzt.

In einem optimalen Testplan würde jeder Test denselben Informationsgehalt besitzen. Wären allerdings zwei Tests absolut identisch, so würde einer von ihnen etwas zum Informationsgehalt des Testplans beitragen, der Informationsgehalt des anderen Tests wäre hingegen gleich Null. Die Reihenfolge in der die identischen Tests ausgewählt werden und die Entscheidung getroffen wird, welcher der beiden Tests einen Informationsgehalt und welcher keinen besitzt ist willkürlich.

5.2.2 Redundanz

Ein Test wird als redundant betrachtet, wenn die durch ihn gewonnene Information kleiner ist als ein gegebener Schwellwert. Einige weitere Eigenschaften des Tests - sowie seine C_{pk} oder der Korrelationswert u.s.w. - können ebenfalls bei der Entscheidung ob dieser Test redundant ist von einem Regel basierten Mechanismus berücksichtigt werden. Dafür wird eine Linearkombination (Wissensfunktion) des Informationsinhaltes und diverser anderer Testparameter verwendet - abhängig von dem analysierten Produkt und Produktionsschritt. Diese Linearkombination verändert das Ranking der Tests derart, dass produktspezifische Zusammenhänge mitberücksichtigt werden können. Der Begriff *redundant* wird hier ausschließlich im informationstheoretischen Sinne verwendet. Genau so, wie starke Korrelation nicht die Abhängigkeit definiert, definiert auch die Redundanz nicht die *Weglassbarkeit*. Aus diesem Grund muss ein Prüfenieur mit entsprechendem Hintergrundwissen die als redundant markierten Tests letztlich als weglassbar bestätigen.

Ist schließlich eine Menge T von Tests analysiert und eine Untermenge $R \subseteq T$ als redundant identifiziert worden, können zwei Testpläne definiert werden. Ein *regulärer* Testplan, der nur die *obligatorischen* Tests $M = T \setminus R$ enthält und ein *erweiterter* Testplan T . Der erweiterte Plan kann bei problematischen bzw. kritischen Bauteilen eingesetzt werden. Ein Hinweis darauf können beispielsweise schlechte oder verdächtige Testergebnisse aus der Menge M oder die Analyse der PCM-Messergebnisse liefern. Der reguläre Testplan M kann in allen anderen Fällen eingesetzt werden und spart dabei die Testzeit und Testkosten der Tests aus R .

Kapitel 6

Experimente

In diesem Kapitel werden die im Rahmen dieser Arbeit durchgeführten Experimentreihen betrachtet. Die erste Reihe befasst sich mit der Prognostizierbarkeit der Wafer-Ausbeute durch die Betrachtung der PCM-Messwerte. Die zweite Experimentreihe untersucht die Möglichkeit des Einsatzes von PCM-Messergebnissen als Indikatoren für „problematische“ Bauteile, um - wie im Abschnitt 5.2.2 vorgestellt - für diese den erweiterten Testplan zu verwenden. Bei allen anderen Bauteilen bzw. Wafers soll der reguläre Testplan verwendet werden um Testkosten einzusparen.

Die Experimente basieren auf den im Abschnitt 4.2 vorgestellten Klassifikationsverfahren. Die Klassifikation geschieht auf Basis historischer Daten ähnlicher Wafer, die für die Initialisierung der Klassifikatoren benötigt werden. Diese Daten werden auch *Trainingsdaten* genannt. Sie beinhalten sowohl IC-Messergebnisse als auch PCM-Messwerte. Anhand der IC-Messergebnisse werden die Wafer in zwei Klassen eingeteilt:

$$Klasse_{(Kriterium \geq \text{Schwellenwert})}, Klasse_{(Kriterium < \text{Schwellenwert})}$$

Die Klassen speichern die PCM-Messwerte zugehöriger Wafer als Beispiele. Während der Klassifizierung werden die PCM-Messergebnisse des zu klassifizierenden Wafers mit den Proben, die in den Klassen gespeichert sind, verglichen. Auf diese Weise kann die Wahrscheinlichkeit für die Zugehörigkeit des zu klassifizierenden Wafers zu einer Klasse bestimmt werden.

Alle folgenden Experimente untersuchen die Genauigkeit der Vorhersage der Klassifikationsalgorithmen, die über die Zugehörigkeit eines Wafers zu einer der obigen Klassen entscheiden. Die berechnete Fehlerrate ist hierbei immer auf ganze Wafer und nicht einzelne Bauteile bezogen. Nähere Informationen zu der Klassifikation und den Klassifikationsverfahren finden sich im Abschnitt 4. Weiterhin werden in beiden Experimentreihen die Kreuzvalidierung (Abschnitt 4.3) und die lineare Diskriminanzanalyse (Abschnitt 4.4) eingesetzt.

6.1 Wafer-Klassifikation

Wie im Abschnitt 3.2 gesehen, besteht der Produktionsprozess aus mehreren Phasen. Gäbe es eine Möglichkeit einen besonders schlechten Wafer früh genug im Prozess zu erkennen, so könnten die später anfallenden Test- und Produktionskosten eingespart werden. Besonders schlechte Wafer lassen sich beispielsweise als solche durch ihre besonders geringe Ausbeute definieren. Würde also eine Möglichkeit existieren die Wafer-Ausbeute in einem frühen Stadium des Prozesses zu bestimmen bzw. vorherzusagen, so könnte über das weitere Vorgehen bezüglich des Wafers bereits zu diesem Zeitpunkt entschieden werden.

Ein solcher Zeitpunkt für die Vorhersage der Ausbeute könnte die 2. Phase sein. Nach der Bestückung der Wafer mit den PCMs werden auf diesen mehrere Tests durchgeführt. Die Ergebnisse dieser Tests (PCM-Messergebnisse) werden zur Analyse der Prognostizierbarkeit der Wafer-Ausbeute verwendet. In mehreren folgenden Experimenten werden alle im Abschnitt 4.2 vorgestellten Klassifikationsverfahren ausprobiert und miteinander verglichen.

Als erstes werden die tatsächlichen Ausbeuten der Wafer beider im Abschnitt 5.1 beschriebener Datensätze in den Diagrammen 6.1 und 6.2 betrachtet.

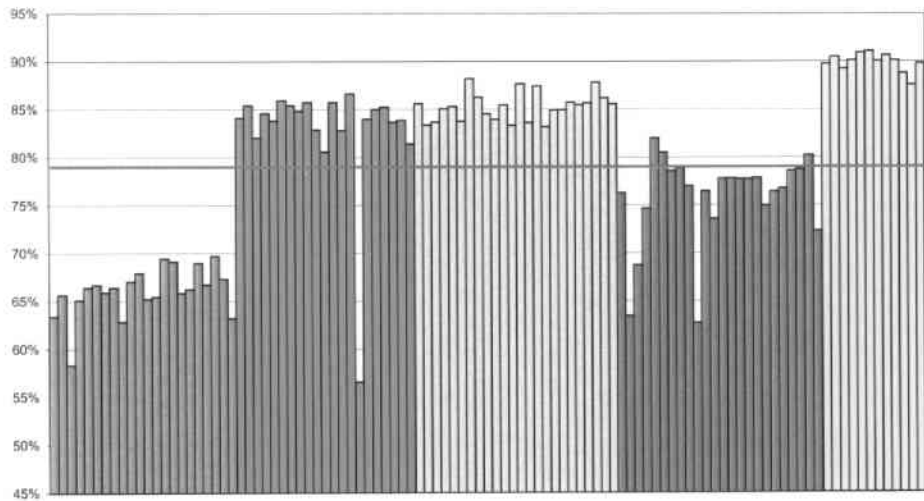


Abbildung 6.1: Ausbeuten des 1. Datensatzes

Unterschiedliche Farben stellen unterschiedliche Lose dar. Die y-Achse repräsentiert die tatsächliche Ausbeute in Prozent.

Es ist zu erkennen, dass einzelne Lose sich untereinander teilweise sehr stark unterscheiden, die Wafer innerhalb eines Loses jedoch meistens eine ähnliche Ausbeute besitzen und sich möglicherweise deshalb auch in weiteren Eigenschaften ähneln.

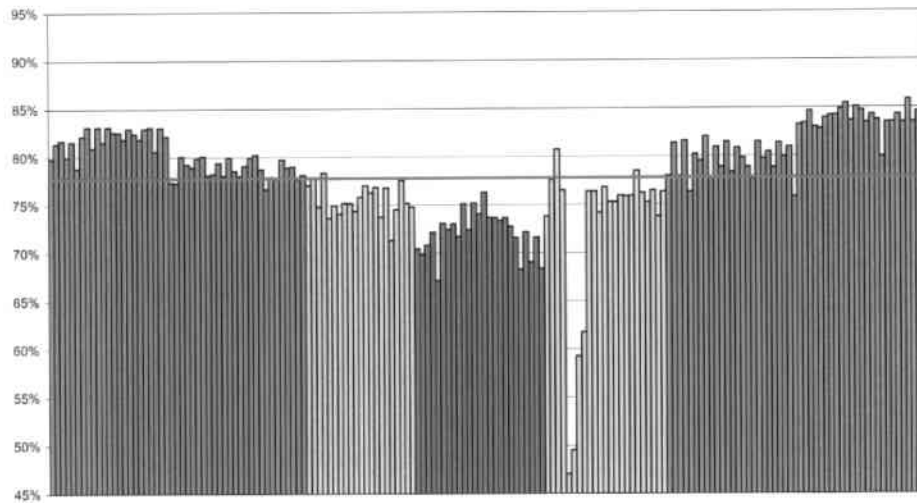


Abbildung 6.2: Ausbeuten des 2. Datensatzes

6.1.1 Klassifikation der Wafer-Ausbeute ohne Verwendung der linearen Diskriminanzanalyse

In diesem Abschnitt wird versucht mit verschiedenen Klassifikationsverfahren die Ausbeute der Wafer beider Datensätze vorherzusagen. Es wird vorerst keine lineare Diskriminanzanalyse verwendet.

Die Fehlerraten der durchgeführten Experimente sind in den Abbildungen 6.3 und 6.4 in Abhängigkeit vom Schwellwert der Ausbeute aufgetragen.

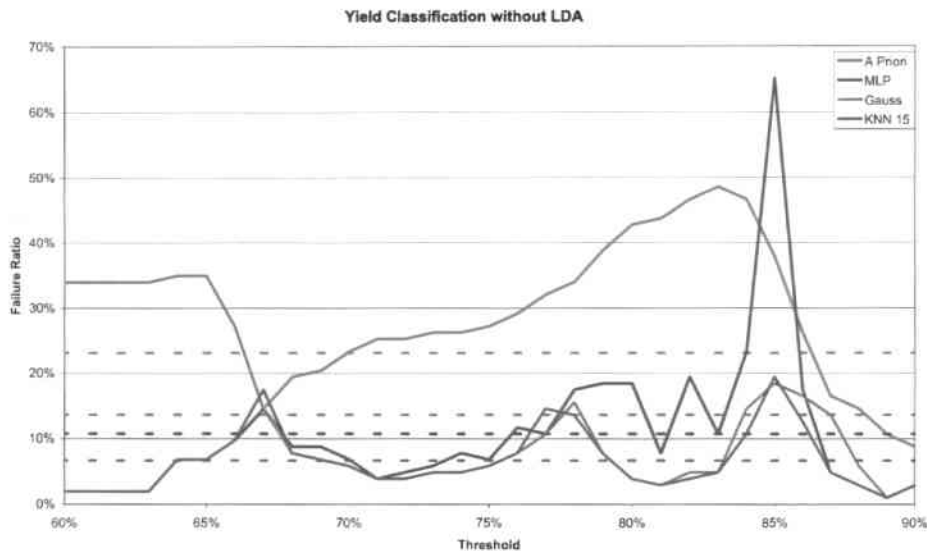


Abbildung 6.3: Fehlerraten: Vorhersagbarkeit der Ausbeuten des 1. Datensatzes ohne Anwendung linearer Diskriminanzanalyse

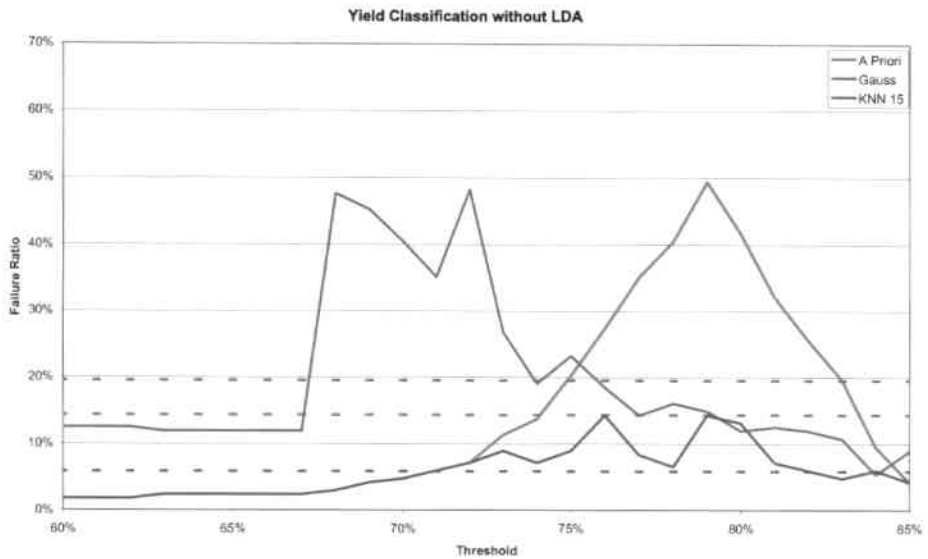


Abbildung 6.4: Fehlerraten: Vorhersagbarkeit der Ausbeuten des 2. Datensatzes ohne Anwendung linearer Diskriminanzanalyse

Der A-Priori-Klassifikator (*A-Priori*), der wie im Abschnitt 4.2.1 vorgestellt keine PCM-Messergebnisse sondern lediglich die a-priori Wahrscheinlichkeiten beider Klassen betrachtet, ist hier, sowie in allen folgenden Grafiken rot dargestellt. Andere Klassifikationsverfahren sollten einen Informationsgewinn (siehe Definition 4.1) dem A-Priori-Klassifikator gegenüber aufweisen, da sie weitere Informationen, wie die erwähnten PCM-Messergebnisse betrachten. Der Informationsgewinn mündet jedoch nicht immer in einer geringeren Fehlerrate, wie auf den Abbildungen zu erkennen ist.

Der einzige in beiden Datensätzen überzeugende Klassifikator ist der Nächste-Nachbar-Klassifikator (*KNN*). Er wird in allen Diagrammen blau dargestellt. Dieser Klassifikator verwendet bei dieser, sowie bei allen anderen Klassifikationen den Wert 15 für k - die Anzahl nächster Nachbarn. Nach der Untersuchung aller für k möglichen Werte im Bereich zwischen 1 und 127, stellte sich heraus, dass für die beiden untersuchten Datensätze der Wert 15 für k bezüglich der Fehlerrate optimal ist. Die durchschnittliche Fehlerquote liegt bei beiden Datensätzen bei ca. 7%, es gibt jedoch mehrere Schwellenwerte, bei welchen die 10% Marke überschritten wird.

Die Ergebnisse des Gauss-Klassifikators (*Gauss*) multivariat und univariat werden in Diagrammen grün dargestellt. Es ist zu erkennen, dass der Klassifikator teilweise große Probleme bei bestimmten Schwellenwerten hat, bei welchen er Fehlerquoten von bis zu 48% erreicht. Beim 2. Datensatz hat er eine durchschnittliche Fehlerrate von ca. 20% und liegt dabei um 6% schlechter als der naive A-Priori-Klassifikator. Wie später zu sehen sein wird, ist dieses Verhalten für den Gauß-Klassifikator auf multidimensionalen Stichproben beider Datensätze typisch. Die hohe Fehlerrate sinkt jedoch signifikant nach der Anwendung der linearen Diskriminanzanalyse.

Der auf Multi-Layer Perzeptronen (*MLP*) aufbauende Klassifikator wird lediglich auf den 1. Datensatz angewendet. Die Problematik dieses Klassifikators liegt in den indeterministischen und zum Teil sehr schlechten Ergebnissen, welche zum Teil deutlich unter denen des A-Priori-Verfahrens liegen. Wie im Abschnitt 4.2.4 beschrieben, wurden mehrere unterschiedliche Konfiguration ausprobiert, die allerdings allesamt keine nennenswerte Verbesserung brachten. Erstaunlicherweise erzielte ein längeres Training mit 1000 Epochen ebenfalls keine Verbesserung. Eine höhere Anzahl an Trainingsepochen wurde nicht getestet, da weder eine Tendenz zur Verbesserung der Fehlerrate festgestellt werden konnte, noch die erforderlichen Kapazitäten zur Verfügung standen. Letztlich kommen mehrere Ursachen für das mangelhafte Klassifikationsverhalten dieses Algorithmus in Frage. Darunter fallen zwar auch zu kleine Trainingszyklen, dennoch kann eine nicht korrekte Dokumentation oder inkorrekte Programmierung der Bibliothek [Rue04] nicht ausgeschlossen werden.

6.1.2 Klassifikation der Wafer-Ausbeute unter Verwendung der linearen Diskriminanzanalyse

In diesem Abschnitt wird auf beiden Datensätzen vor der eigentlichen Klassifikation zuerst die lineare Diskriminanzanalyse (siehe Abschnitt 4.4) durchgeführt. Die Klassifikationsergebnisse beider Datensätze sind auf den Abbildungen 6.5 und 6.6 zu sehen. Die Ordinate stellt wieder die Fehlerrate und die Abszisse die verschiedenen Schwellenwerte für die Ausbeute in Prozent dar.

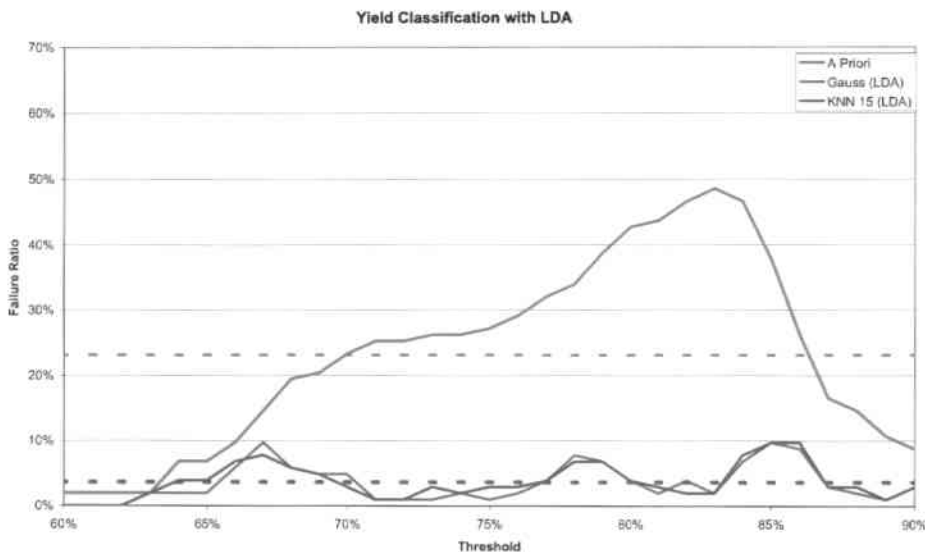


Abbildung 6.5: Fehlerraten: Vorhersagbarkeit der Ausbeuten des 1. Datensatzes nach Anwendung linearer Diskriminanzanalyse

Der Unterschied zwischen diesen Ergebnissen und den Ergebnissen aus dem Abschnitt 6.1.1 ist deutlich zu erkennen. Weder der Nächste-Nachbar- noch der Gauss-Algorithmus erreichen hier eine Fehlerquote von über 10%.

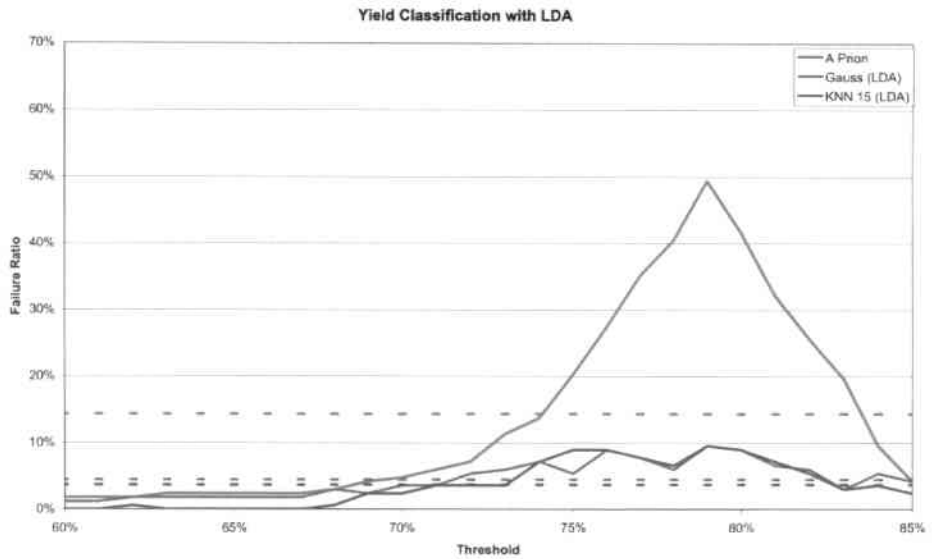


Abbildung 6.6: Fehlerraten: Vorhersagbarkeit der Ausbeuten des 2. Datensatzes nach Anwendung linearer Diskriminanzanalyse

Die durchschnittlichen Fehlerraten liegen jetzt zwischen 4 und 5 Prozent. Bei einigen Schwellenwerten sinkt die Fehlerrate sogar auf unter 2%. Es ist weiterhin zu erkennen, dass beide Verfahren trotz völlig unterschiedlicher Annahmen und Funktionsweise zum Teil sehr ähnliche, wenn nicht gar gleiche Ergebnisse erzielen. Ebenfalls ist der beachtliche Vorsprung „intelligenter“ Klassifikatoren gegenüber dem A-Priori-Verfahren klar auszumachen.

6.1.3 Precision-Recall Auswertung der Wafer-Ausbeute Klassifikation

Zur eingehenden Untersuchung der im letzten Abschnitt durchgeführten Vorhersagen sollen im Folgenden die Precision-Recall Statistiken der Klassifikatoren für die beiden Klassen $Klasse_{(Kriterium \geq Schwellenwert)}$ und $Klasse_{(Kriterium < Schwellenwert)}$ betrachtet werden. Die Vorhersagen aus dem Abschnitt 6.1.1 werden hier nicht aufgeführt, da sie durchgehend unter den Ergebnissen des letzten Abschnittes liegen und keine praxisrelevanten Resultate liefern.

Der *Recall* beschreibt die Vollständigkeit einer Klassifikation. Er ist definiert als der Anteil erkannter, zur $Klasse_i$ gehörender Wafer an den insgesamt zur $Klasse_i$ gehörenden Wafer:

$$Recall(Klasse_i) = \frac{\{Wafer \in Klasse_i\} \cap \{Erkannte Wafer\}}{\{Wafer \in Klasse_i\}}$$

Die *Precision* beschreibt die Genauigkeit einer Klassifikation. Sie ist definiert als der Anteil erkannten, zur $Klasse_i$ gehörender Wafer von allen bei der Klassifikation zur $Klasse_i$ klassifizierten Wafer:

$$Precision(Klasse_i) = \frac{\{Wafer \in Klasse_i\} \cap \{Erkannte Wafer\}}{\{Erkannte Wafer\}}$$

Die Abbildungen 6.7 bis 6.10 zeigen die Precision-Recall Ergebnisse der beiden Datensätze.

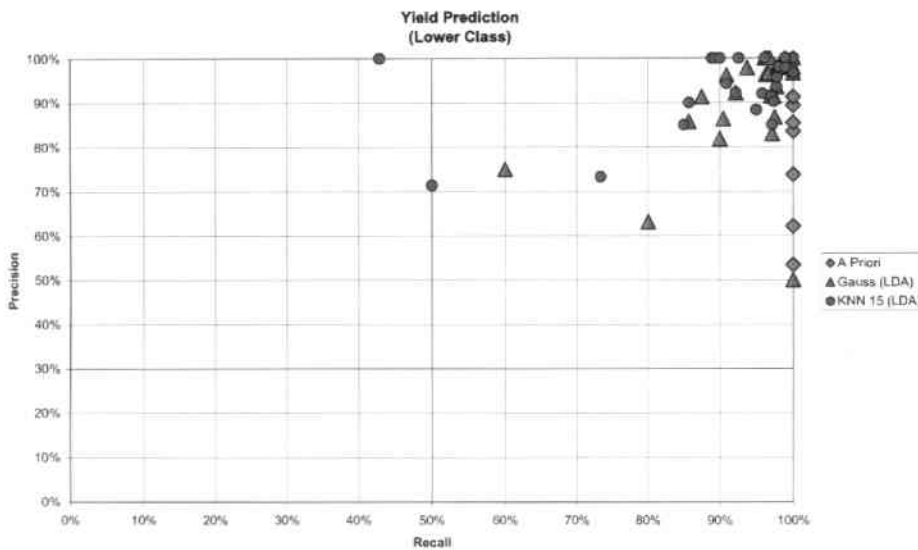


Abbildung 6.7: Precision-Recall Statistik für die Vorhersage der Ausbeute des 1. Datensatzes für die $Klasse_{(Kriterium < Schwellenwert)}$ unter Verwendung der linearen Diskriminanzanalyse

Die roten Datenpunkte repräsentieren auch hier den A-Priori-Klassifikator, die grünen Punkte stehen für das Gauss-Verfahren und die blauen für den Nächste-Nachbar-Algorithmus. Auf der Ordinate sind die Precision- und auf der Abszisse die Recall-Werte dargestellt.

6.1.4 Fazit: Wafer-Klassifikation

In den beiden letzten Abschnitten wurde gezeigt, dass eine Vorhersage der Wafer-Ausbeute beider vorliegender Datensätze auf einem Fehlerniveau von ca. 5% prinzipiell möglich ist. Ebenfalls zu sehen war eine signifikante Verbesserung der Klassifikationsergebnisse nach der Anwendung linearer Diskriminanzanalyse.

Die ohne linearer Diskriminanzanalyse ausgeführte Klassifikationen konnten mit keinem verwendeten Algorithmus die für reale Einsatzzwecke nötige Fehlerfreiheit erzielen. Die Fehlerraten des Nächste-Nachbar-Verfahrens lagen

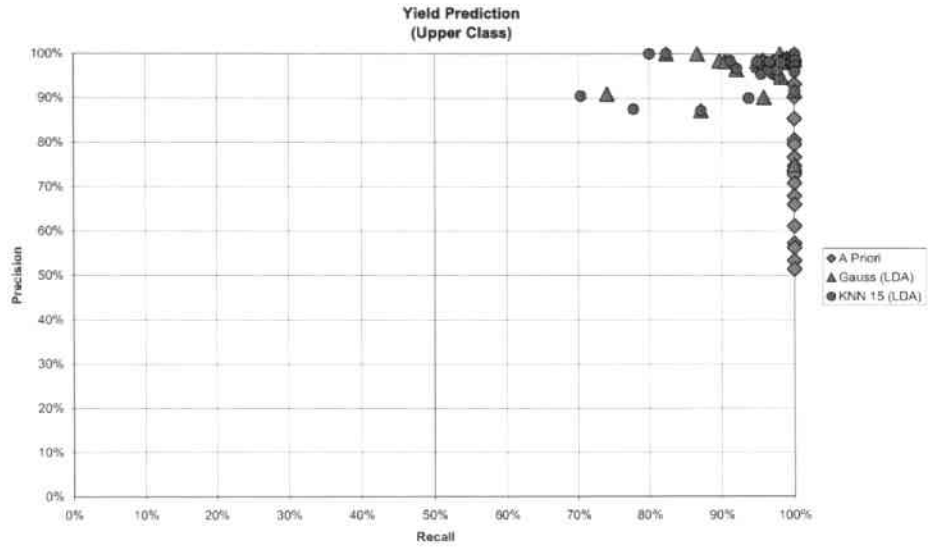


Abbildung 6.8: Precision-Recall Statistik für die Vorhersage der Ausbeute des 1. Datensatzes für die $Klasse_{(Kriterium \geq Schwel\text{lenwert})}$ unter Verwendung der linearen Diskriminanzanalyse

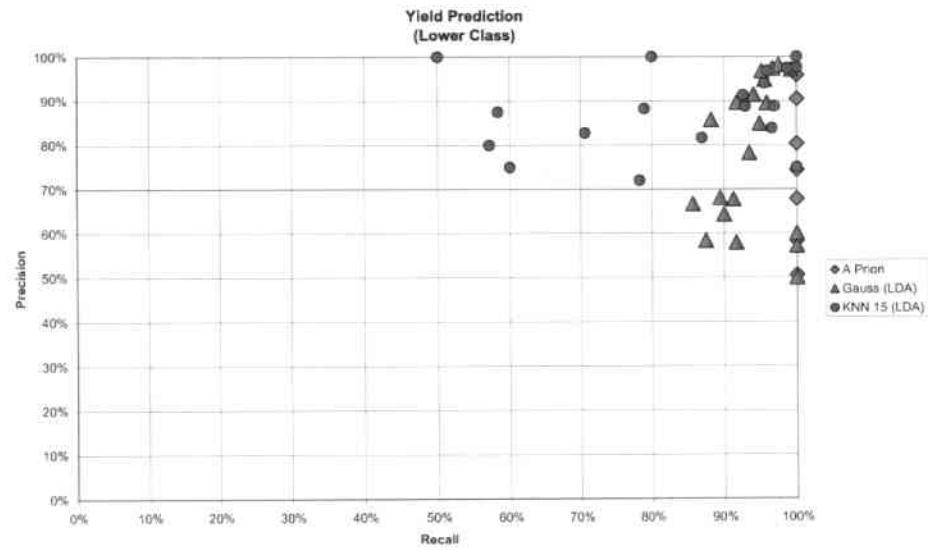


Abbildung 6.9: Precision-Recall Statistik für die Vorhersage der Ausbeute des 2. Datensatzes für die $Klasse_{(Kriterium < Schwel\text{lenwert})}$ unter Verwendung der linearen Diskriminanzanalyse

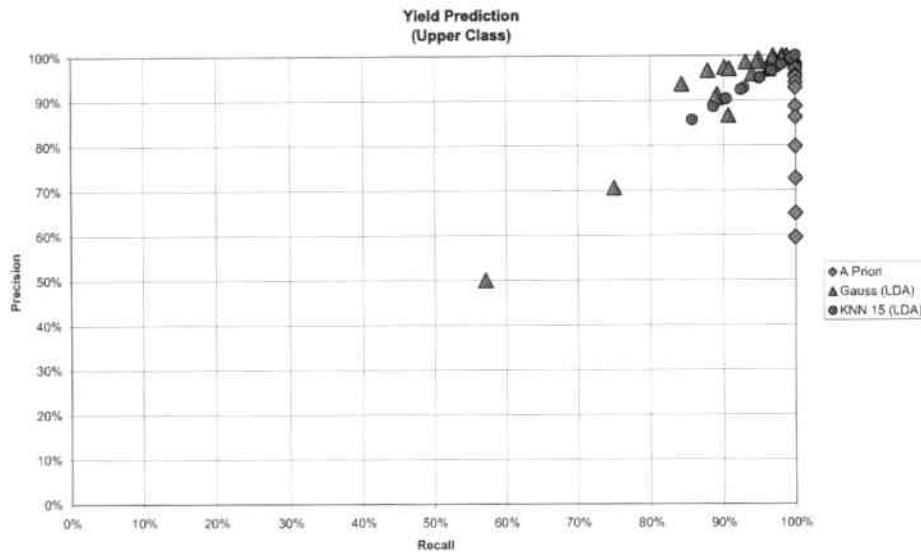


Abbildung 6.10: Precision-Recall Statistik für die Vorhersage der Ausbeute des 2. Datensatzes für die Klasse ($K_{\text{kriterium}} \geq \text{Schwellenwert}$) unter Verwendung der linearen Diskriminanzanalyse

zwar im Durchschnitt zwischen 7 und 8 Prozent, doch erreichten die Ausschläge vor allem in den Ausbeute-Bereichen mit größerer Dichte Werte zwischen 14 und 19 Prozent.

Mit dem Einsatz der linearen Diskriminanzanalyse veränderte sich das Bild und die Fehlerraten gingen stark zurück. Dieses Verhalten lässt sich auch in weiteren, folgenden Experimenten beobachten. Die durchschnittliche Fehlerrate des Gauss-Klassifikators sank bei der Untersuchung der Ausbeute des 2. Datensatzes um ganze 14%. Die Anwendung dieses Klassifikators scheint generell nur nach einer vorhergehenden Diskriminanzanalyse sinnvoll zu sein.

6.2 Testplan-Klassifikation

Im Abschnitt 5.2 wurde eine Möglichkeit gezeigt den allgemeinen Wafer-Testplan in einen redundanten und einen obligatorischen Teil aufzuspalten. Dadurch entstehen zwei Testpläne: Ein regulärer, verkürzter Testplan, ohne Redundanz und ein erweiterter Testplan, mit allen Testplänen. Der reguläre Testplan kann im Normalfall bei allen Bauteilen bzw. Wafern verwendet werden. Der erweiterte Testplan kommt nur dann zum Einsatz, wenn aufgrund der Auswertung bestimmter Testergebnisse aus dem regulären Testplan oder aufgrund bestimmter Eigenschaften der PCM-Messwerte darauf geschlossen werden kann, dass Problemfälle vorliegen.

Im Folgenden werden zuerst Experimente zur Untersuchung der Vorhersagbarkeit einzelner IC-Tests und anschließend Experimente zur Analyse der Pro-

gnostizierbarkeit von mehreren zusammengefassten IC-Tests zu Testsätzen betrachtet. Die errechnete Fehlerrate bezieht sich dabei immer auf ganze Wafer. Befindet sich also beispielsweise auf dem gesamten Wafer nur *ein* Bauteil, dessen Testwerte außerhalb der Toleranzgrenzen liegen und wird dieser Wafer trotzdem als fehlerfrei klassifiziert, so soll der *komplette* Wafer als falsch vorhergesagt gelten.

6.2.1 Einzeltestvorhersage

In diesem Abschnitt wird die Genauigkeit untersucht, mit der sich einzelne IC-Testergebnisse nach der Untersuchung der PCM-Messwerte vorhersagen lassen. Die Klassenaufteilung für die Klassifikatoren erfolgt nach folgender Regel:

- $Klasse_{(Kriterium < Schwellenwert)}$ soll die PCM-Messwerte aller Wafer enthalten, auf denen ein bestimmter Test t_i *keine* Werte gefunden hat, die außerhalb des Toleranzbereichs liegen.
- $Klasse_{(Kriterium \geq Schwellenwert)}$ hingegen erhält die PCM-Messwerte aller Wafer, bei welchen t_i *mindestens einen* Wert außerhalb des Toleranzbereichs festgestellt hat.

Somit ist $Kriterium$ gleich der Anzahl Werte außerhalb des Toleranzbereichs, $Schwellenwert = 0,5$ und es gilt:

$$\begin{aligned} Klasse_{(Kriterium < Schwellenwert)} &= Klasse_{(keine Defekte)}, \\ Klasse_{(Kriterium \geq Schwellenwert)} &= Klasse_{(mindestens ein Defekt)} \end{aligned}$$

Einzeltestvorhersage ohne Verwendung der linearen Diskriminanzanalyse

Gegenstand der Klassifizierung auf den Abbildungen 6.11 und 6.12 sind die einzelnen Tests, die auf ihre Prognostizierbarkeit mit den bekannten Klassifikatoren untersucht werden. Die lineare Diskriminanzanalyse kam dabei nicht zum Einsatz. Auf den beiden Grafiken ist auf der y-Achse die Fehlerrate der Klassifikatoren dargestellt und die x-Achse repräsentiert die einzelnen vorhergesagten Tests.

Wie bereits im Abschnitt 6.1 gesehen, ist die durchschnittliche Fehlerrate des Nächste-Nachbar-Algorithmus etwas besser als die des A-Priori-Klassifikators. Das Gauss-Verfahren scheint dagegen bei beiden Datensätzen große Probleme zu haben. Die Fehlerraten dieses Verfahrens liegen bei vielen Tests weit über der 40% Marke. Die durchschnittlichen Fehlerquoten des Gauss-Klassifikators liegen zwischen 29 und 38 Prozent, die des Nächste-Nachbar-Klassifikators zwischen 10 und 16 Prozent. Mit diesen Fehlerraten könnte keines der Verfahren in einer realen Produktionsumgebung eingesetzt werden.

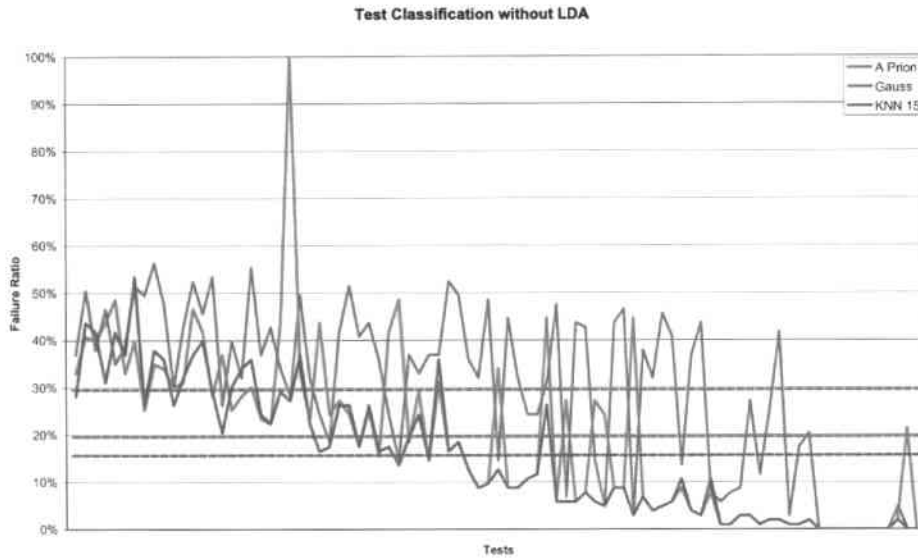


Abbildung 6.11: Fehlerraten: Klassifizierung aller 87 IC-Tests des 1. Datensatzes ohne Verwendung der linearen Diskriminanzanalyse

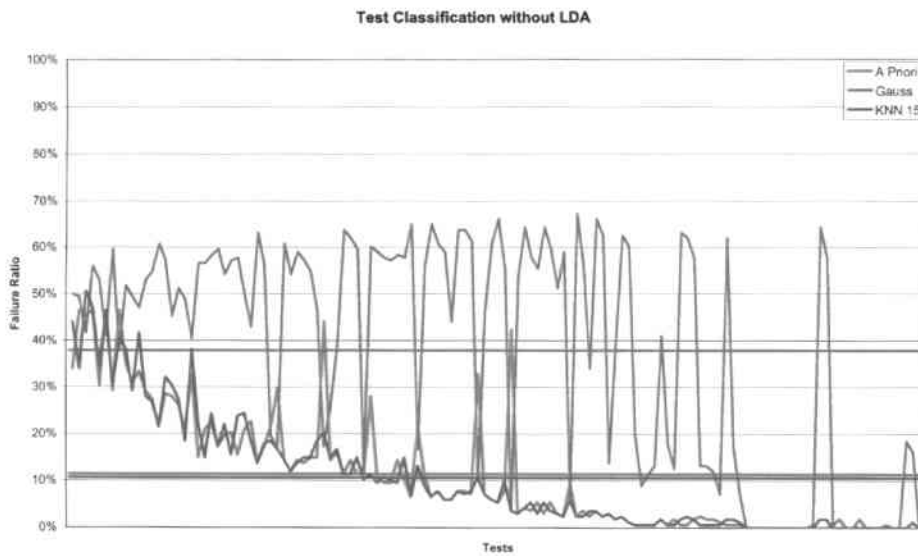


Abbildung 6.12: Fehlerraten: Klassifizierung aller 129 IC-Tests des 2. Datensatzes ohne Verwendung der linearen Diskriminanzanalyse

Einzeltestvorhersage unter Verwendung der linearen Diskriminanzanalyse

Im Abschnitt 6.1.2 zeigte sich, dass der Einsatz der linearen Diskriminanzanalyse große Wirkung auf die Ergebnisse haben kann. Deshalb soll nun in den Diagrammen 6.13 und 6.14 die Vorhersagbarkeit einzelner Tests nach einer vorangehenden Diskriminanzanalyse betrachtet werden.

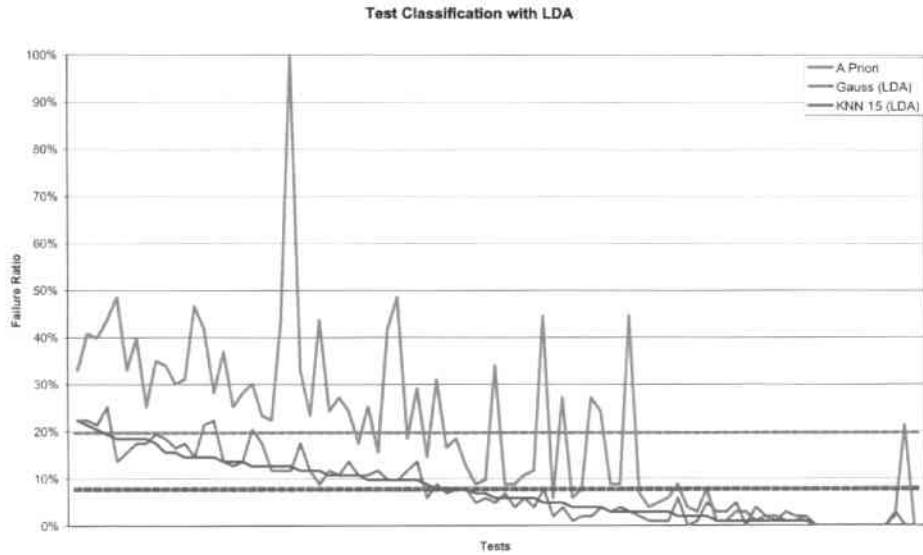


Abbildung 6.13: Fehlerraten: Klassifizierung aller 87 IC-Tests des 1. Datensatzes unter Verwendung der linearen Diskriminanzanalyse

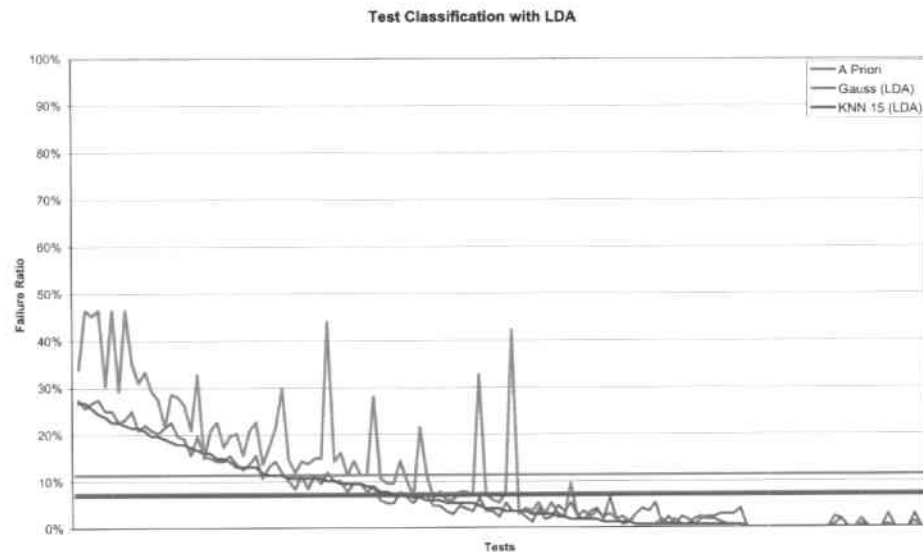


Abbildung 6.14: Fehlerraten: Klassifizierung aller 129 IC-Tests des 2. Datensatzes unter Verwendung der linearen Diskriminanzanalyse

Wie erwartet, ist wieder eine grosse Differenz zwischen den Klassifikationsergebnissen mit und ohne linearer Diskriminanzanalyse zu beobachten. Der Gauss-Klassifikator scheint auch dieses Mal nach einer durchgeführten Diskriminanzanalyse nicht mehr Schwierigkeiten zu haben als der Nächste-Nachbar-Algorithmus. Die durchschnittliche Fehlerquote beider Klassifikatoren liegt nun bei 8 bis 9 Prozent. Ein großer Anteil der Tests lässt sich sogar mit einer Fehlerrate von unter 5% vorhersagen, wenngleich der tatsächliche Informationsgewinn (siehe Definition 4.1) nur selten hoch ist.

Precision-Recall Auswertung der Einzeltestvorhersage

In den Abbildungen 6.15 bis 6.18 sollen wieder, wie im Abschnitt 6.1.3 die Precision-Recall Statistiken der Klassifikatoren nach durchgeführter linearer Diskriminanzanalyse betrachtet werden. Erneut sind die Statistiken nach Datensätzen und den beiden Klassen aufgeteilt. Die Vorhersagen ohne Anwendung der linearen Diskriminanzanalyse werden hier nicht aufgeführt.

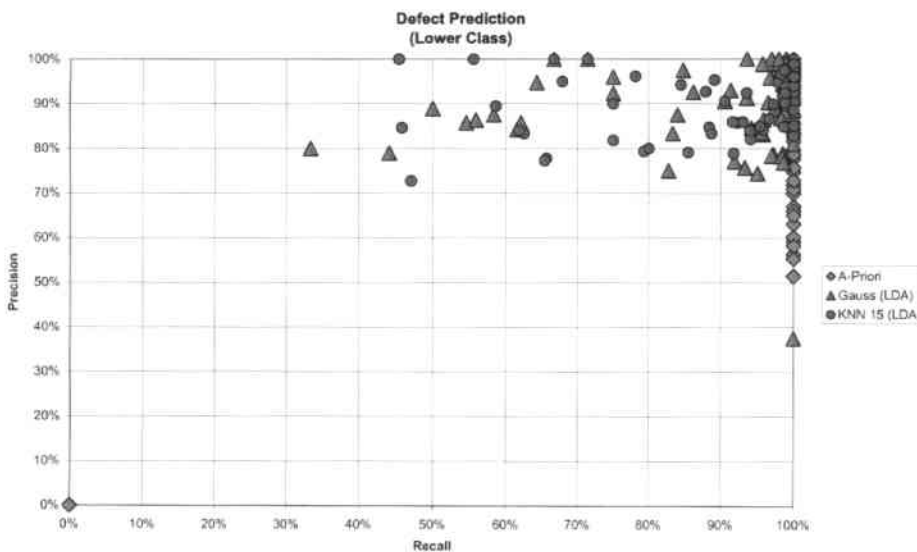


Abbildung 6.15: Precision-Recall Statistik für die Vorhersage der Defekte des 1. Datensatzes für die Klasse $(keine\ Defekte)$ unter Verwendung der linearen Diskriminanzanalyse

Die roten Punkte repräsentieren den A-Priori-Klassifikator, die grünen Datenpunkte den Gauss-Klassifikator und die blauen den Nächste-Nachbar-Algorithmus. Auf der y-Achse sind die Precision- und auf der x-Achse die Recall-Werte aufgetragen.

Fazit: Einzeltestvorhersage

In den letzten Abschnitten war wieder erkennbar, dass sich ein bedeutender Anteil ($> 40\%$) der Tests beider Datensätze mit einer annehmbaren Fehlerwahrscheinlichkeit von 5% vorhersagen lässt. Ebenfalls verdeutlicht wurde

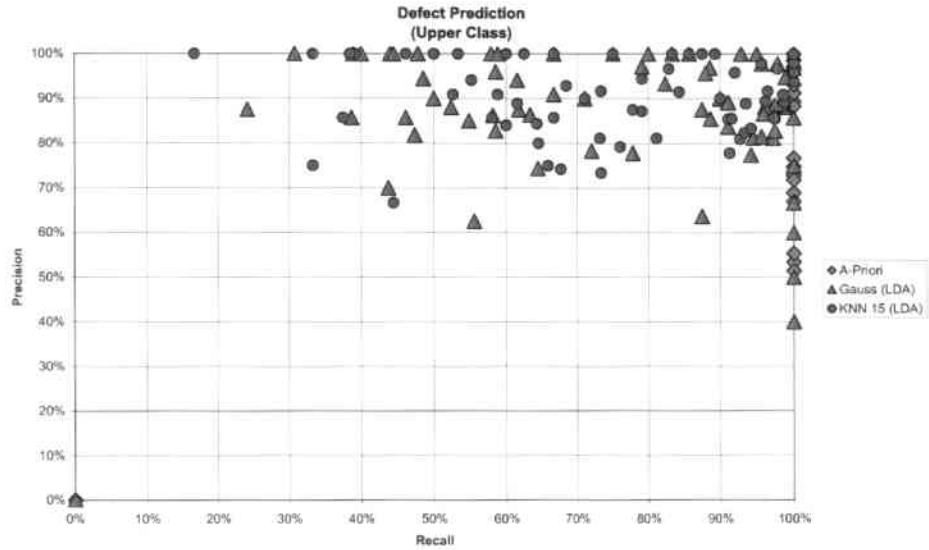


Abbildung 6.16: Precision-Recall Statistik für die Vorhersage der Defekte des 1. Datensatzes für die *Klasse*_(mindestens ein Defekt) unter Verwendung der linearen Diskriminanzanalyse

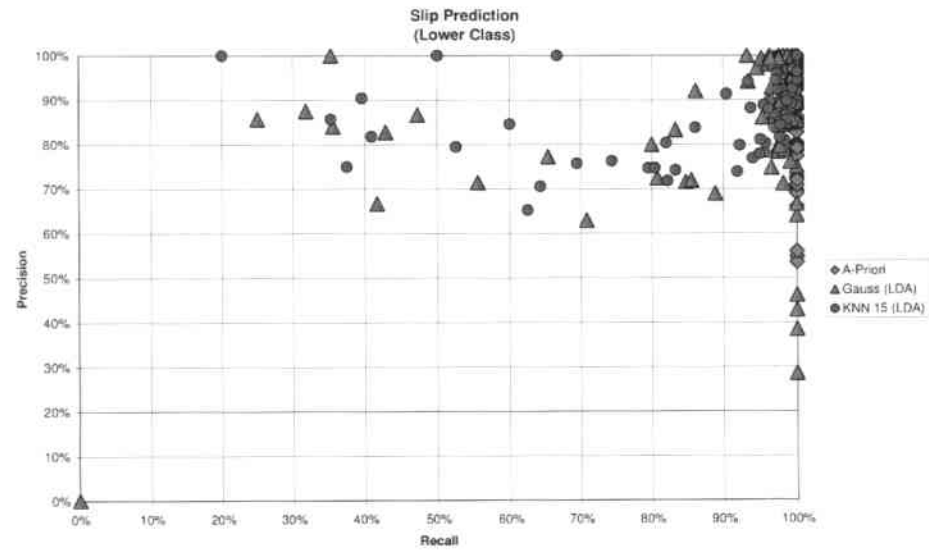


Abbildung 6.17: Precision-Recall Statistik für die Vorhersage der Defekte des 2. Datensatzes für die *Klasse*_(keine Defekte) unter Verwendung der linearen Diskriminanzanalyse

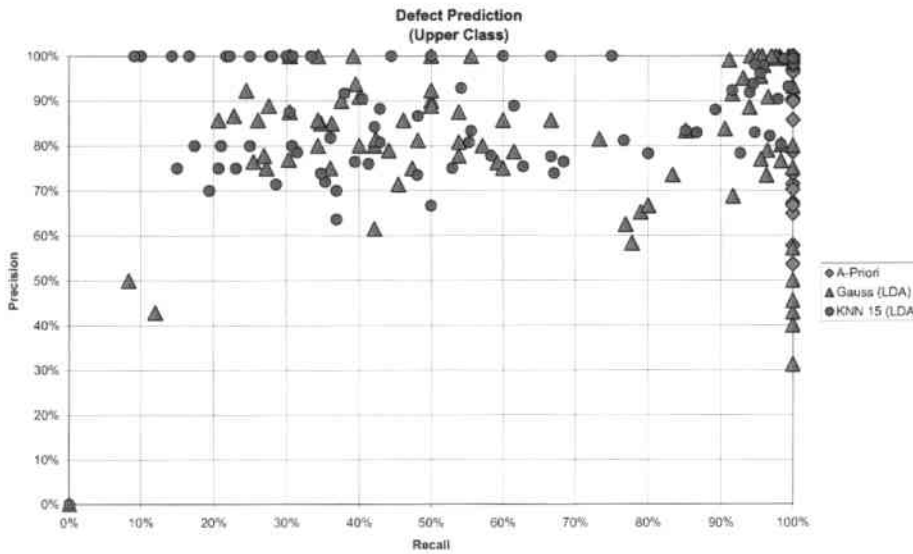


Abbildung 6.18: Precision-Recall Statistik für die Vorhersage der Defekte des 2. Datensatzes für die Klasse $(\text{mindestens ein Defekt})$ unter Verwendung der linearen Diskriminanzanalyse

die Wichtigkeit des Einsatzes der linearen Diskriminanzanalyse zur Senkung durchschnittlicher Fehlerraten auf ein angemessenes Niveau.

6.2.2 Testplanauswahl

Nachdem in den letzten Abschnitten die Genauigkeit der Prognose einzelner IC-Testergebnisse untersucht wurde, werden jetzt die Fehlerquoten der Vorhersage ganzer Testsätze betrachtet. Ein Testsatz T_i besteht aus der Testmenge $\{t_1, t_2, \dots, t_k\}$ und meldet immer dann ein defektes Bauteil, wenn mindestens ein Test t_1, t_2, \dots, t_k ein solches Bauteil entdeckt. Somit gilt:

$$T_i(\text{Wafer}_w) = \sum_{t_t \in T_i} t_t(\text{Wafer}_w)$$

Mit dem Software-Tool reduTec wurden aufgrund der Wahl verschiedener Parameter für beide Datensätze jeweils 12 potentiell redundante Testsätze mit jeweils 1 bis 21 Tests für den erweiterten Testplan bestimmt. Um diese Tests sicher weglassen zu können, wird ein robustes Verfahren benötigt um problematische Bauteile sicher zu erkennen. Bei diesen problematischen Chips würde der reguläre Testplan, wie im Abschnitt 5.2.2 vorgestellt, um die prinzipiell redundanten Tests erweitert werden. Die Genauigkeit der drei Klassifikationsverfahren (A-Priori, Gauss und Nächste-Nachbar) bei der Identifikation problematische Bauteile wird in folgenden Abschnitten analysiert.

Die Klassifikation erfolgt nach einer ähnlichen Regel, wie bereits im Abschnitt 6.2.1:

- $Klasse_{(Kriterium < Schwellenwert)}$ soll die PCM-Messwerte aller Wafer enthalten, auf denen ein bestimmter Testsatz $T_i = \{t_1, t_2, \dots, t_k\}$ keine Werte gefunden hat, die außerhalb des Toleranzbereichs liegen.
- $Klasse_{(Kriterium \geq Schwellenwert)}$ erhält die PCM-Messwerte aller Wafer, bei welchen der Testsatz $T_i = \{t_1, t_2, \dots, t_k\}$ mindestens einen Wert außerhalb des Toleranzbereichs festgestellt hat.

Damit lässt sich das *Kriterium* gleich der Anzahl der Werte außerhalb des Toleranzbereichs setzen bei einem *Schwellenwert* = 0, 5. Somit ergibt sich wieder:

$$\begin{aligned} Klasse_{(Kriterium < Schwellenwert)} &= Klasse_{(keine Defekte)} \\ Klasse_{(Kriterium \geq Schwellenwert)} &= Klasse_{(mindestens ein Defekt)}. \end{aligned}$$

In den vergangenen Abschnitten 6.1 und 6.2.1 wurde deutlich, dass eine sinnvolle Klassifikation nur nach einer vorhergehenden Diskriminanzanalyse durchgeführt werden kann. Aus diesem Grund wird im Folgenden auf die Klassifikationen ohne der Anwendung einer Diskriminanzanalyse verzichtet.

Klassifikation der Testpläne unter Verwendung der linearen Diskriminanzanalyse

Mit Hilfe der linearen Diskriminanzanalyse und der Verwendung bekannter Klassifikationsverfahren (A-Priori, Gauss und Nächste-Nachbar) soll hier analysiert werden, mit welcher Fehlerquote es möglich ist die Resultate des erweiterten Testplans zu prognostizieren. Zu diesem Zweck werden wieder (siehe Abschnitt 6.2.1) die PCM-Messwerte eines Wafers analysiert. Anhand dieser geschieht die Zuordnung zu einer der beiden oben erwähnten Klassen und somit auch die Vorhersage über die Resultate der IC-Tests des redundanten Testplans. In den Diagrammen 6.19 und 6.20 werden die Fehlerquoten der Vorhersage auf Basis der PCM-Messwerte untersucht.

Trotz des deutlichen Informationsgewinns im ersten Datensatz sind die Fehlerraten (im Durchschnitt ca. 13%) teilweise zu hoch, im Gegensatz dazu fällt der Informationsgewinn (siehe Definition 4.1) beim zweiten Datensatz moderat aus, doch auch die durchschnittliche Fehlerrate scheint sich bei 5% einzupendeln. Da der Informationsgewinn relativ hoch ist und die Tests ohnehin als potentiell redundant bestimmt wurden, stellt sich hiermit die Frage nach dem Ursprung der unerwartet hohen Fehlerrate. Die Antwort darauf könnte die vorher eingeführte Grobkörnigkeit bei der Untersuchung der Genauigkeit einer Klassifikation sein. Es werden stets nur ganze Wafer gezählt und die Fehlklassifikation eines einzelnen Chips wird als eine falsche Vorhersage des ganzen Wafers gewertet. Zur endgültigen Klärung dieser Frage und Untermauerung der aufgestellten These werden deshalb im nächsten Abschnitt die unentdeckten Defekte (Fehlklassifikationen) vorhergehender Prognose betrachtet.

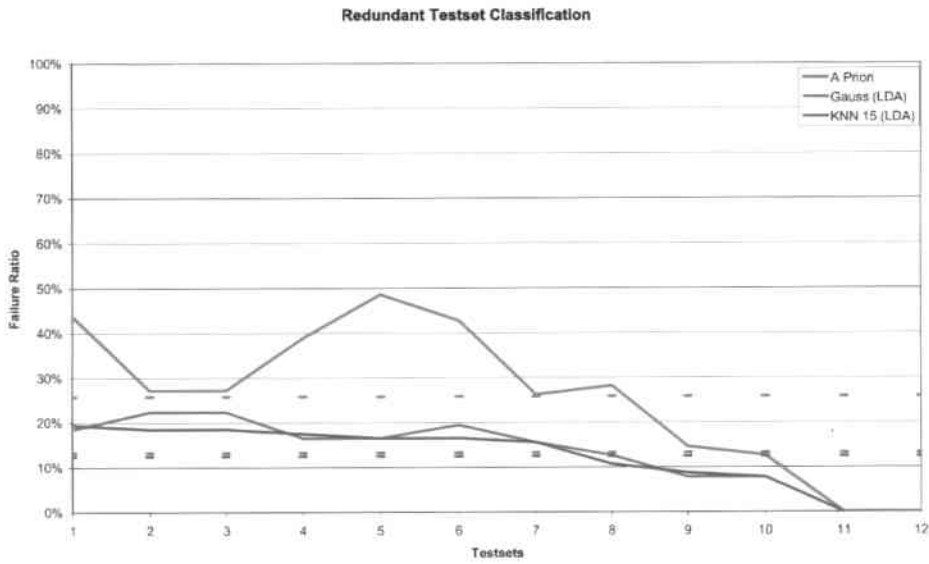


Abbildung 6.19: Fehlerraten: Klassifizierung der 12 potentiell redundanten Testsätze des 1. Datensatzes nach Anwendung der linearen Diskriminanzanalyse

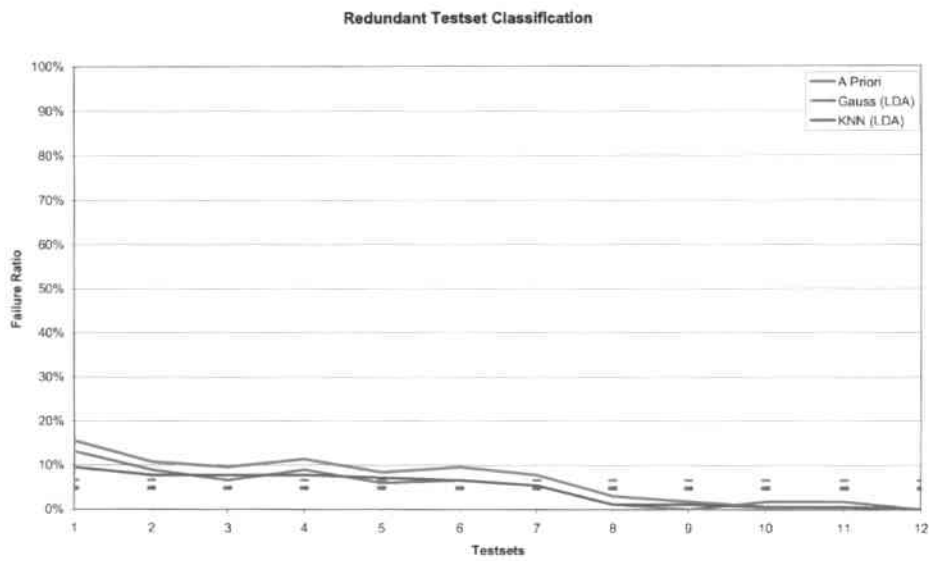


Abbildung 6.20: Fehlerraten: Klassifizierung der 12 potentiell redundanten Testsätze des 2. Datensatzes nach Anwendung der linearen Diskriminanzanalyse

Betrachtung der bei der Klassifikation unentdeckter Defekte

Wie in obigen Abschnitten definiert, erfolgt die Ermittlung der Fehlerraten nur Wafer-bezogen. Das heißt, dass ein einzelnes unentdecktes, defektes Bauteil (*Escape*) als fehlerhafte Klassifikation eines gesamten Wafers betrachtet wird. Daher erscheint es sinnvoll in den folgenden zwei Abbildungen 6.21 und 6.22 die gesamte Anzahl unentdeckter, defekter Bauteile für jeden prinzipiell redundanten Testsatz der Datensätze 1 und 2 aufzuführen.

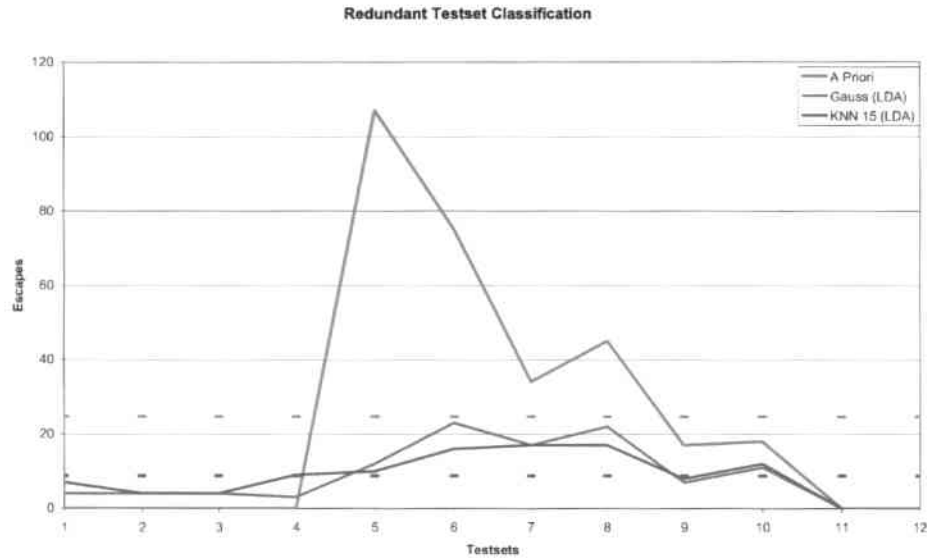


Abbildung 6.21: Summe unentdeckter Defekte je Testsatz bei der Klassifikation der 12 möglicherweise redundanten Testsätze auf 103 Wafers des 1. Datensatzes unter Verwendung der linearen Diskriminanzanalyse

Auf den beiden Abbildungen ist deutlich zu sehen, dass trotz der etwas erhöhten Fehlerrate im vorigen Abschnitt, die Anzahl der Escapes im Durchschnitt - und vor allem pro Wafer gerechnet - relativ gering ist. Das wiederum bedeutet, dass die im obigen Abschnitt diskutierte Fehlerrate entweder auf *false positives* oder auf eine nur sehr geringe Menge unentdeckter, defekter Chips pro Wafer zurückzuführen ist. Diese könnten sogar aufgrund versehentlicher Kratzer oder stellenweiser Verunreinigungen entstanden sein und ließen somit nicht anhand der PCM-Daten detektieren.

Precision-Recall Auswertung der Testplan-Klassifikation

In den Grafiken 6.23 bis 6.26 werden die Precision-Recall Statistiken der Klassifikationen aus den beiden vorigen Abschnitten betrachtet.

6.3 Kostenrechnung

In diesem Abschnitt werden die durch den Einsatz der Klassifikationsverfahren ermöglichte Ersparnisse betrachtet. Als Rechnungsgrundlage werden die tatsächlichen Zahlen bzw. Kosten der Firma Atmel verwendet. Demnach kostet

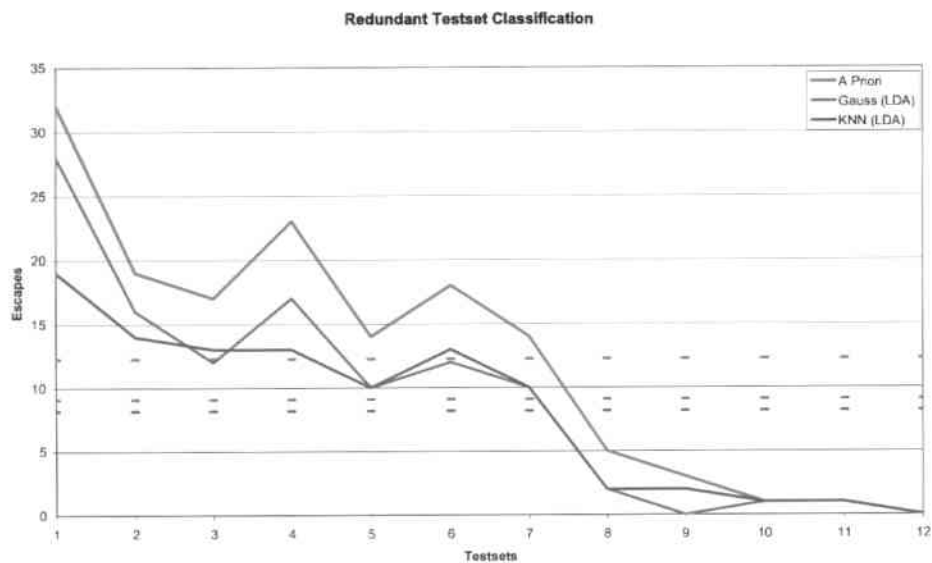


Abbildung 6.22: Summe unentdeckter Defekte je Testsatz bei der Klassifikation der 12 möglicherweise redundanten Testsätze auf 168 Wafern des 2. Datensatzes unter Verwendung der linearen Diskriminanzanalyse

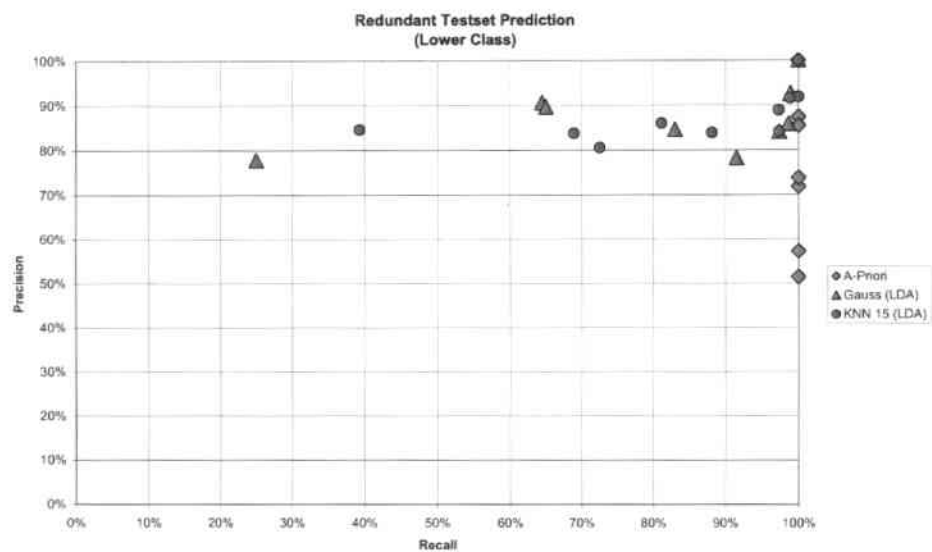


Abbildung 6.23: Precision-Recall Statistik für die Vorhersage der Defekte der potentiell redundanten Testsätze des 1. Datensatzes für die Klasse_(keine Defekte) unter Verwendung der linearen Diskriminanzanalyse

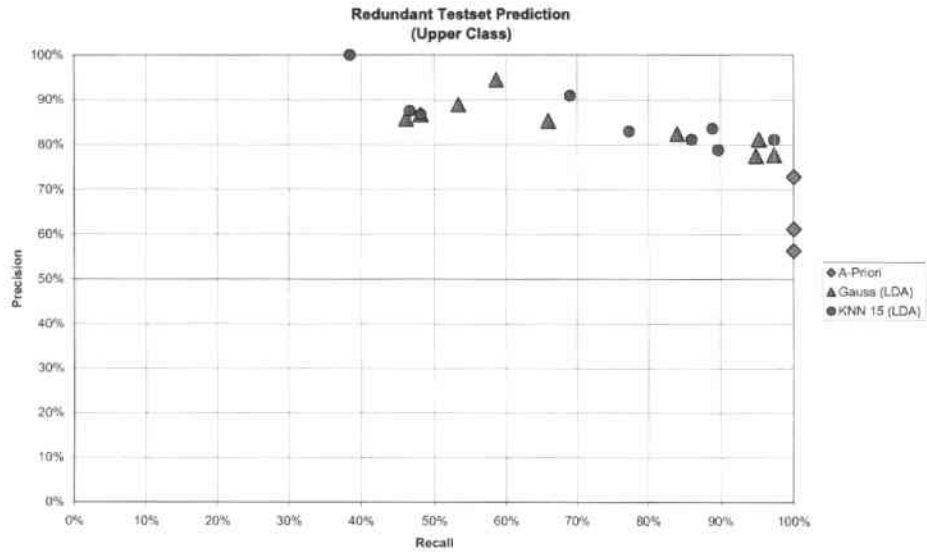


Abbildung 6.24: Precision-Recall Statistik für die Vorhersage der Defekte der potentiell redundanten Testsätze des 1. Datensatzes für die $Klasse_{(mindestens\ ein\ Defekt)}$ unter Verwendung der linearen Diskriminanzanalyse

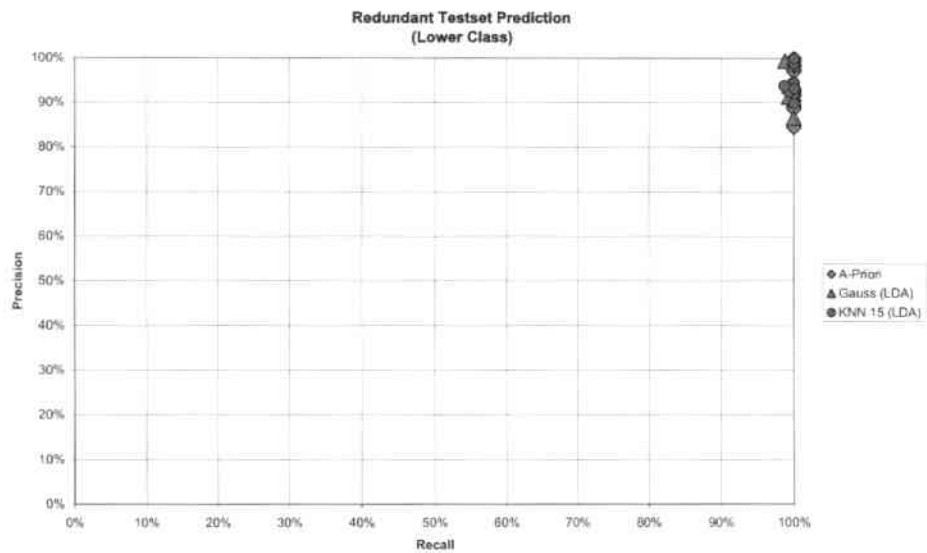


Abbildung 6.25: Precision-Recall Statistik für die Vorhersage der Defekte der potentiell redundanten Testsätze des 2. Datensatzes für die $Klasse_{(keine\ Defekte)}$ unter Verwendung der linearen Diskriminanzanalyse

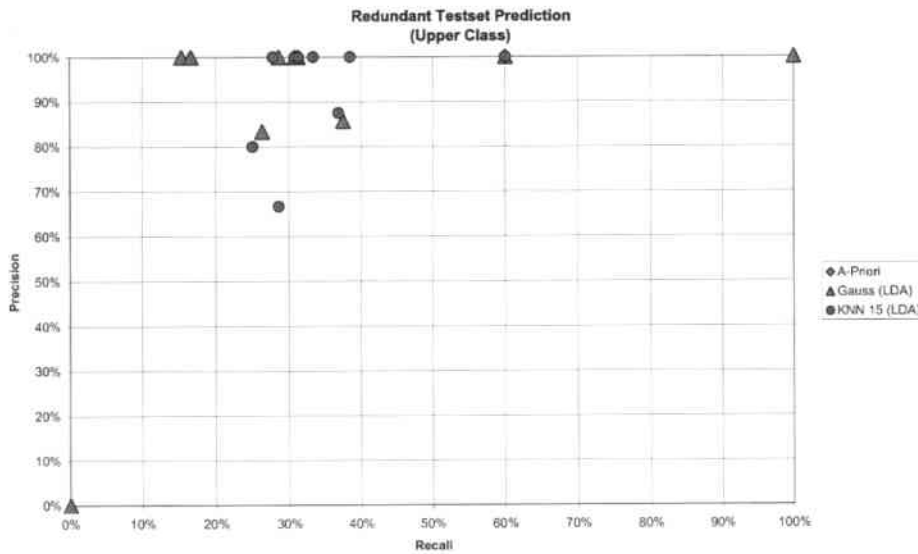


Abbildung 6.26: Precision-Recall Statistik für die Vorhersage der Defekte der potentiell redundanten Testsätze des 2. Datensatzes für die Klasse (mindestens ein Defekt) unter Verwendung der linearen Diskriminanzanalyse

ein Test auf einem IC eines Wafers ca. 0,0001 EUR. Ein unentdecktes, defektes IC (Escape) verursacht hingegen durch unnötiges Einpacken und finales Testen Kosten in Höhe von bis zu 1 EUR. Man berechnet den Gewinn als:

$$\text{Gewinn} = \left[\text{Anzahl}(\text{Wafer}_{\text{korrekt erkannt}}) \cdot \frac{\text{Kosten}}{\text{Wafer}} \right] - \left[\text{Anzahl}(\text{Escapes}) \cdot \frac{\text{Kosten}}{1 \text{ Escape}} \right] \tag{6.1}$$

und

$$\frac{\text{Kosten}}{\text{Wafer}} = \frac{\text{Anzahl}(\text{ICs})}{1 \text{ Wafer}} \cdot \frac{\text{Anzahl}(\text{Tests})}{1 \text{ IC}}$$

Wobei $\text{Anzahl}(\text{Wafer}_{\text{korrekt erkannt}})$ nur die Anzahl der richtig klassifizierten Wafer angibt.

Auf den Abbildungen 6.27 und 6.28 sind die insgesamt erzielbaren Gewinne auf den beiden Datensätzen dargestellt, wenn die 12 prinzipiell redundanten Testsätze aus den vorigen Abschnitten verwendet werden. In den beiden folgenden Grafiken 6.29 und 6.30 wird hingegen die Ersparnis pro Wafer des Datensatzes angegeben.

Wie bereits im Abschnitt 5.1 gesehen, liegen die Testkosten der 3. Phase für einen Wafer zwischen 28 EUR und 33 EUR. Nimmt man die Spitzenwerte möglicher Ersparnisse pro Wafer, so errechnet sich demnach eine Kostensenkung pro Wafer von bis zu 8,2% beim ersten und 13,6% beim zweiten Datensatz.

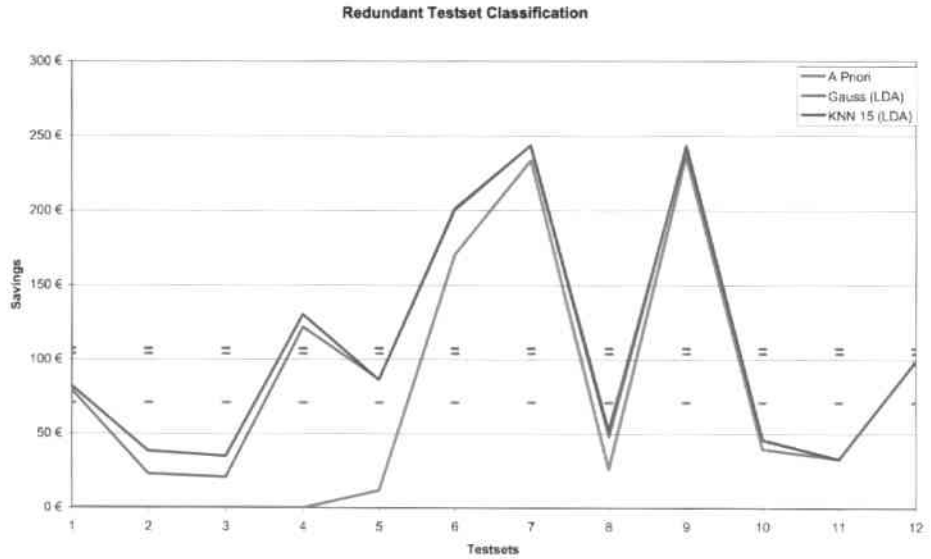


Abbildung 6.27: Gewinn: Durch Gleichung 6.1 berechnete Ersparnis bei der Klassifikation der 12 möglicherweise redundanten Testsätze auf 103 Wafern des 1. Datensatzes nach Anwendung linearer Diskriminanzanalyse

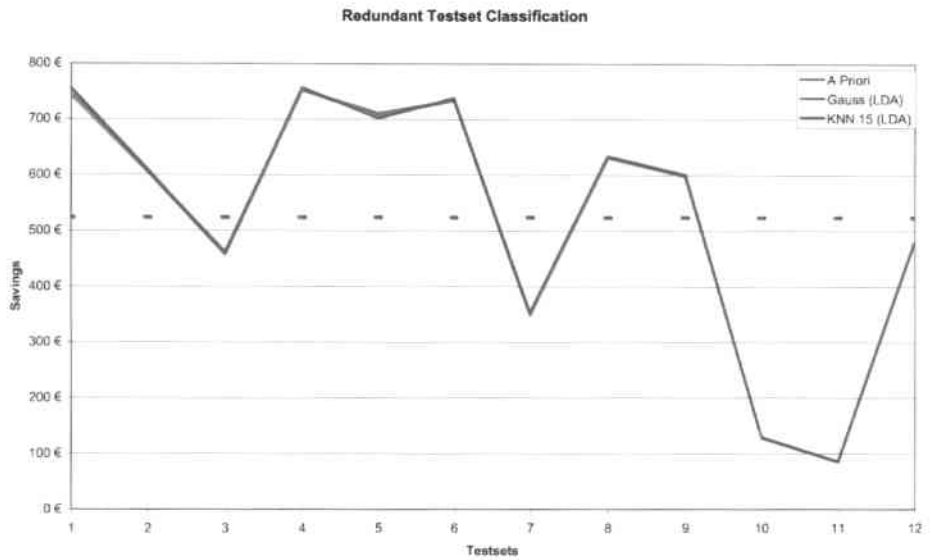


Abbildung 6.28: Gewinn: Durch Gleichung 6.1 berechnete Ersparnis bei der Klassifikation der 12 möglicherweise redundanten Testsätze auf 168 Wafern des 2. Datensatzes nach Anwendung linearer Diskriminanzanalyse)

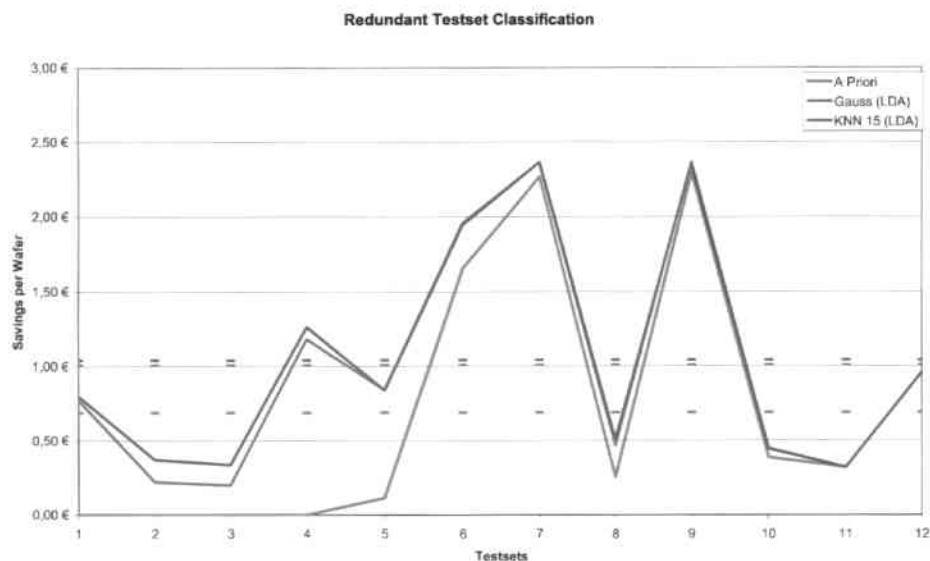


Abbildung 6.29: Gewinn: Durch Gleichung 6.1 berechnete Ersparnis bei der Klassifikation der 12 möglicherweise redundanten Testsätze pro Wafer des 1. Datensatzes unter Verwendung der linearen Diskriminanzanalyse

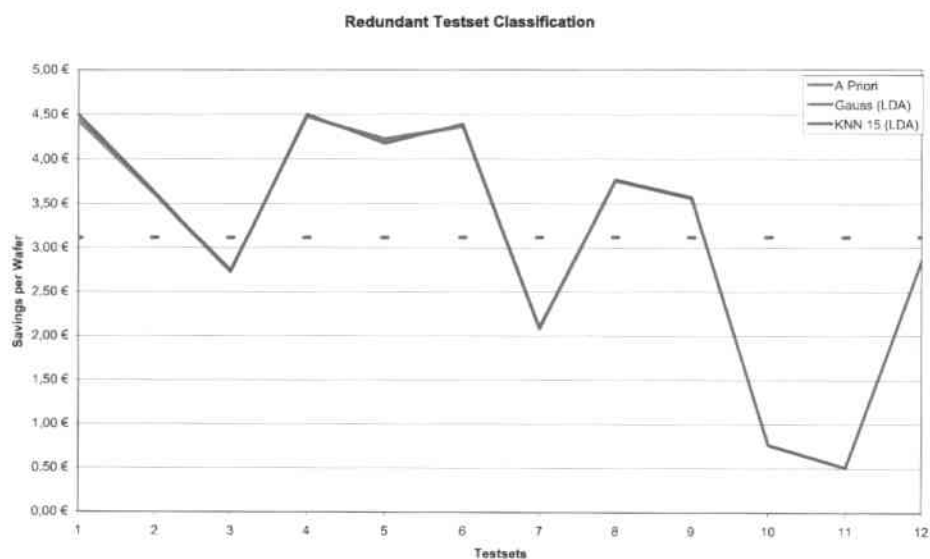


Abbildung 6.30: Gewinn: Durch Gleichung 6.1 berechnete Ersparnis bei der Klassifikation der 12 möglicherweise redundanten Testsätze pro Wafer des 2. Datensatzes unter Verwendung der linearen Diskriminanzanalyse

6.4 Betrachtung der Gültigkeit

Abschließend soll die Gültigkeit der in den vorhergehenden Abschnitten erzielten Ergebnisse betrachtet und ihre möglichen Einschränkungen genannt werden.

Mögliche Einschränkungen der internen Gültigkeit:

- Eine mögliche Einschränkung stellt die Verwendung der Kreuzvalidierung dar. Durch den Einsatz dieses Verfahrens gab es in mindestens einer Klasse stets genug Beispiele, die aus dem gleichen Los stammten, wie das zu klassifizierende Wafer. Da die Abbildungen 6.1 und 6.2 die Vermutung verstärken, dass innerhalb eines Loses die Wafer sich auf eine bestimmte Art ähneln, könnten die Ergebnisse verfälscht worden sein.
- Für die Experimente wurden bekanntlich die von der Firma Atmel zur Verfügung gestellten Daten verwendet. Deshalb kann nicht sichergestellt werden, dass die Auswahl dieser streng zufällig geschehen ist. Demnach besteht die Möglichkeit, dass die Auswahl der Wafer bestimmte nicht abschätzbare Auswirkungen auf die Klassifikationsergebnisse hatte.

Mögliche Einschränkungen der externen Gültigkeit:

- Die beiden untersuchten Datensätze bestehen aus Wafern zweier unterschiedlicher Produkte, die von der Firma Atmel ausgesucht wurden. Die vorgenommene Auswahl muss jedoch nicht zwangsläufig alle anderen Wafer dieser Produkte repräsentieren.
- Die analysierten Produkte sind nicht unbedingt für die gesamte Produktpalette der Firma Atmel repräsentativ.
- Die Produkte der Firma Atmel sind eventuell für andere Firmen der Halbleiterbranche nicht repräsentativ.

Kapitel 7

Ausblick

In dieser Arbeit wurden Möglichkeiten aufgezeigt wie mit Hilfe gängiger Klassifikationsverfahren auf Basis von PCM-Messergebnissen aus dem zweiten Produktionsschritt die IC-Messergebnisse der nächsten Produktionsstufe zumindest teilweise vorhergesagt werden können. Ein weiterer Schritt wäre die Vorhersage der IC-Messergebnisse einzelner Wafer-Regionen aufgrund der Lokalisierung bestimmter PCM-Messwerte.

Denkbar ist ebenfalls, dass aufgrund der Analyse der PCM-Messwerte nur ganz bestimmte Messungen auf den ICs durchgeführt werden. Der Unterschied zu dem heutigen Verfahren ist dabei, dass für jeden einzelnen IC ein *eigener* Testplan zusammengestellt werden würde - und zwar in Abhängigkeit von den Messergebnissen der vorhergehenden Phase. Diese könnten sich dann wiederum von einer Wafer-Region zur anderen unterscheiden.

Ein weiterer wichtiger Aspekt ist, dass das in dieser Arbeit vorgestellte Verfahren ohne größere Modifikation auch auf andere Produktionsphasen - als nur die zweite und die dritte - angewendet werden könnte. Vor allem interessant wären in diesem Hinblick die finalen Tests, die letztlich über die Auslieferung eines Chips zum Kunden entscheiden.

Letztlich könnten die in dieser Arbeit eingesetzten Algorithmen auch als eine Art vollautomatischer und dynamischer Erweiterung des in [RKG05] beschriebenen Verfahrens eingesetzt werden.

Literaturverzeichnis

- [Cha01] **M. J. Chandra.** *Statistical Quality Control.* CRC Press LLC, 2001.
- [DHS00] **R. Duda, P. Hart und D. Stork.** *Pattern Classification (2nd Edition).* Wiley-Interscience, October 2000.
- [Fuk90] **K. Fukunaga.** *Introduction to Statistical Pattern Recognition (2nd Edition).* Academic Press, Inc., September 1990.
- [ITR03] **ITRS.** The ITRS International Technology Roadmap for Semiconductors. online: <http://public.itrs.net/>, 2003.
- [Mil98] **L. Milor.** A Tutorial Introduction to Research on Analog and Mixed-Signal Circuit Testing. *IEEE Transactions on Circuits and Systems-II*, 45(10):1389–1407, Oktober 1998.
- [RKG05] **I. Rogina, G. Karner und W. Gehring.** Dynamic and Online Control of Flying Probe and Wafer Testers, Report 2005/01. Technical report, optimiSE GmbH, Haid-und-Neu Str. 7, 76131 Karlsruhe, Deutschland, Januar 2005.
- [Rog04] **I. Rogina.** Exploiting Redundancy Information in Logs from IC and PCM Tests. In *Proceedings of the 5th European AEC/APC Conference*, April 2004, Dresden.
- [Rue04] **C. Rueegg.** NeuroBox. online: <http://neurobox.sourceforge.net/>, 2004.
- [RvSK05] **I. Rogina, H.-M. von Staudt und G. Karner.** Controlling Test Plans by Information-Content-Based Redundancy Analysis. In *eingereicht zum 11th Annual International Mixed-Signals Testing Workshop, IMSTW'05*, Juni 2005, Cannes.