

Klassifikation metallischer Reststoffe in Echtzeit

Alexandra Walter

&



gestellt von Kai Krycki

Dieses Problem wurde bearbeitet von Carolin Biener, Friedrich Kurth, Robert Ribstein, John Andermann, Hala Ahmad und Felix Haag.

Kurzzusammenfassung

Unter geeigneten Bedingungen können Metalle unendlich oft recycelt werden. Um Qualitätsverluste zu vermeiden, muss die Zusammensetzung der Metallegierungen vor dem Recyceln untersucht werden. Durch Gammastrahlung kann das gelingen. Dazu wird eine Probe mit Neutronen bestrahlt und die Wechselwirkung zwischen dem Strahl und der Probe gemessen.

1 Problemstellung



Abbildung 1: Metallische Reststoffe

Wie können die Kurzzeitmessungen den bekannten Langzeitmessungen zugeordnet werden? Gibt es gemeinsame „Features“, die in den Kurzzeit- und Langzeitspektren identifiziert werden können? Wie können diese automatisch erkannt werden? Können bei den bekannten Langzeitspektren Gruppen gebildet werden, über die eine grobe Vorabzuordnung stattfinden kann? Kann darauf aufbauend eine noch schnellere, genauere Zuordnung stattfinden?

2 Daten

Gegeben waren genaue Langzeitmessungen von zwanzig bekannten Stoffgemischen. Zusätzlich 250 Kurzzeitmessungen unbekannter Gemische, welche zuzuordnen waren. Langzeitmessungen entsprechen Messungen von etwa 1-4 Stunden und die Kurzzeitmessungen entsprechen wenigen Sekunden.

3 Physikalischer Hintergrund

3.1 PGNAA allgemein

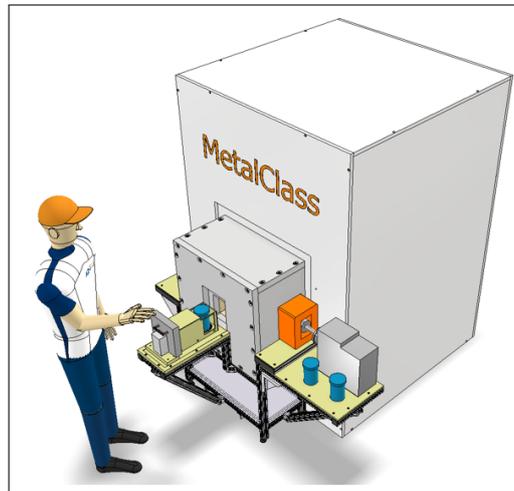


Abbildung 2: Scannen

PGNAA steht für ‚Prompt Gamma Neutron Activation Analysis‘ und ist eine Zerstörungsfreie, berührungslose Analyse Technik auf nuklearer Basis. Es kann auf verschieden große Mengen angewendet werden und kann Proben in bis zu 50 Zentimeter Tiefe erfassen. Sie ist verzögerungsfrei und der Bereich der gegebenen Analyse-Methode beginnt bei 100-150 keV. Es ist möglich eine Auflösung von 4000 bis 32000 ‚channels‘ zu bekommen.

3.2 Aufbau des Geräts

Es gibt einen kleinen Teilchenbeschleuniger welcher Schweres Wasser (Deuterium) auf atomarer Ebene beschleunigt und mit anderem schwerem Wasser fusioniert. Dadurch entsteht Wasserstoff3 (H3) und freie Neutronen. Diese werden dann auf die Atomkerne der unbekanntnen Probe geschossen. Die resultierenden Gamma Strahlen werden dann von einen reinen Germaniumkristall aufgenommen der Verhältnisse mäßig groß ist. Der Germanium-Block befindet sich in einen vakuumierten Raum der durch eine elektronische Kühlung auf einer konstanten Temperatur gehalten wird. Durch eine Halterung aus Aluminium wird er befestigt. Die ersten, gemessenen Werte haben wenig Aussagekraft, da die Schutzhülle des Detektors schwach-energetische Photonen nicht durchlässt. Der Germanium-Block wandelt die Gamma-Strahlen in elektrische Impulse um. Die Hülle besteht aus Graphit. Es kann durch die Graphit Hülle zu ungenaueren Messdaten kommen und neben den ‚Peaks‘ kann es auch zu Rauschen im Graphen der Messdaten kommen.

3.3 Reaktion und Gamma-Strahlung

Bei der Erzeugung der Gamma-Strahlen auf Subatomarer Ebene handelt es sich um eine subatomare Reaktion mit niederenergetischen oder thermischen Neutronen. Die gesamte Energie des Neutrons wird auf den Atomkern der Probe übertragen. Dies aktiviert den Nukleus des Atoms. Das energetisch angeregte Atom wird zum Isotop und hat daraus folgend eine radioaktive Emission. Diese Emissionen sind sekundäre prompte Gammastrahlen. Die Isotope eines Atomkerns haben stoffspezifisch individuelle Energien. Da die ausgesonderten Strahlungen unterschiedliche Energie haben und somit isotoop-spezifische Wellen haben, ist es möglich die verschiedenen Stoffe mit dem Detektor zu unterscheiden. Der Block Germanium wandelt die radioaktive Strahlung in elektrische Impulse.

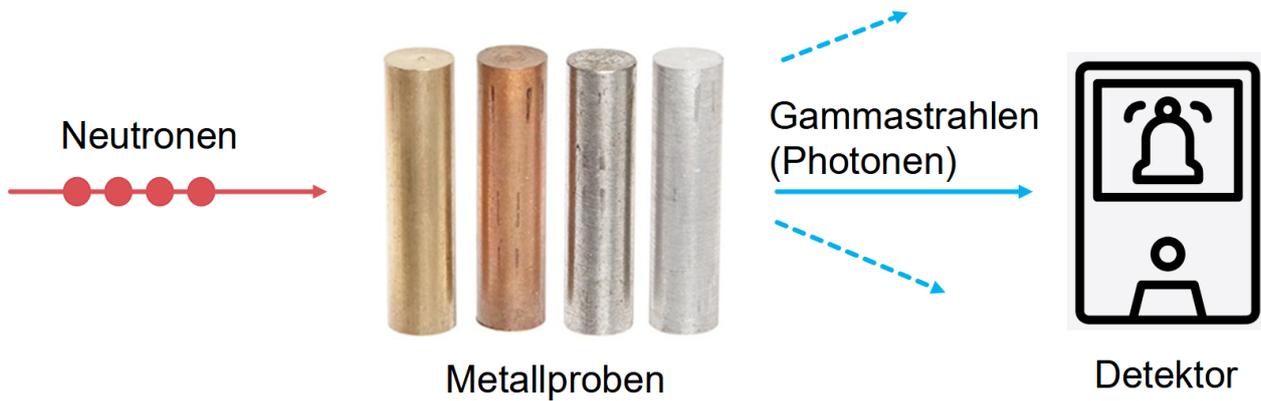


Abbildung 3: Neutronenbeschuss

4 Vorgehensweise

4.1 Problem überblicken und vereinfachen

Nachdem uns die Projekte vorgestellt wurden und unserer Gruppe eins davon zugeteilt wurden, sind wir gemeinsam mit Alexandra Walter, unserer Betreuerin, und Simone Hillemacher in unseren Gruppenraum gegangen, wo die Arbeit sofort losging.

Zuerst setzten wir uns mit der Problemstellung auseinander und tauschten uns über unsere ersten Eindrücke aus. Um ein besseres Gefühl für unsere Daten und das Problem zu entwickeln, versuchten wir erst, die Datensätze zu visualisieren, indem wir sie in Graphen plotten. Dieser Ansatz scheiterte jedoch daran, dass der Rechner die große Menge an Datenpunkten nicht graphisch darstellen konnte.

Daraufhin versuchten wir das Problem zu vereinfachen, indem wir die Menge an Datenpunkten reduzieren. Hierzu sammelten wir Ideen und kamen dabei auf einen Ansatz, den wir jedoch schnell wieder verwerfen mussten. Der Plan, zum Visualisieren die Menge der Datenpunkte zu reduzieren, indem wir von hunderten Päckchen den Durchschnitt berechnen und zu einem Wert zusammenfassen, barg das Risiko, dass wir die höchsten Punkte, die „Peaks“, nicht erfassen können. Den Computer auf den größten und kleinsten Wert in einem Hunderterpäckchen prüfen zu lassen, war eine weitere Möglichkeit. Doch auch da bestand die Gefahr, dass wir dadurch innerhalb der Päckchen wichtige Punkte nicht erfassen.

Die Datenmenge zu reduzieren war für den Lösungsansatz, den wir kurze Zeit später entwickelten, ohnehin unwichtig, weil dieser keine Bestimmung der Peaks durch visuelle Analyse mehr erforderte.

4.2 Vergleich über Produktbildung

Dieser Ansatz bestand im Prinzip darin, die Datensätze der gegebenen Materialkategorien channelweise mit dem einer unbekanntenen Probe zu multiplizieren und die Produkte zu summieren. Diese Summen sollten größer sein, wenn die Peaks aufeinander liegen.

$$\sum_{i=0}^C M_i \cdot P_i$$

Bei dieser Methode gibt es jedoch die Herausforderung, dass die Counts nicht nur von dem Material abhängig sind, sondern vor allem von der Messdauer. Graphisch ausgedrückt liegen die Graphen der Vergleichsmaterialien auf verschiedenen Höhen. Dieses Verfahren hätte höchstwahrscheinlich jede Probe dem gleichen Material zugeordnet.

Es war uns also klar, dass die Datensätze der Materialien normiert werden mussten. Unsere Methode hierfür war es, den Anteil jedes einzelnen Channels an den Gesamtcounts zu berechnen.

$$M'_n = \frac{M_n}{\sum_{i=0}^C M_i}$$

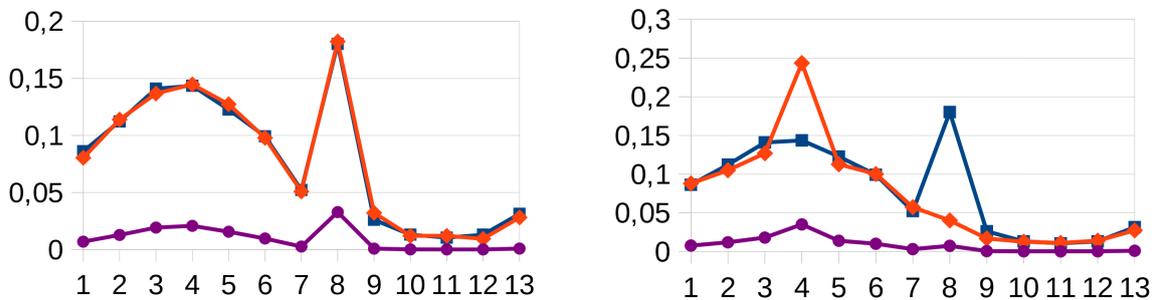


Abbildung 4: Beispieldiagramme für die channelweise Multiplikation

Somit stand der erste Ansatz, die Proben zuzuordnen. Die Gruppe teilte sich auf. Wer programmieren konnte, arbeitete an der Umsetzung in Python mithilfe der Entwicklungsoberfläche "Jupyter", während die Anderen mit unserer Betreuerin die Grundlagen des Programmierens mit Python erarbeiteten.

4.3 Steigungsansatz

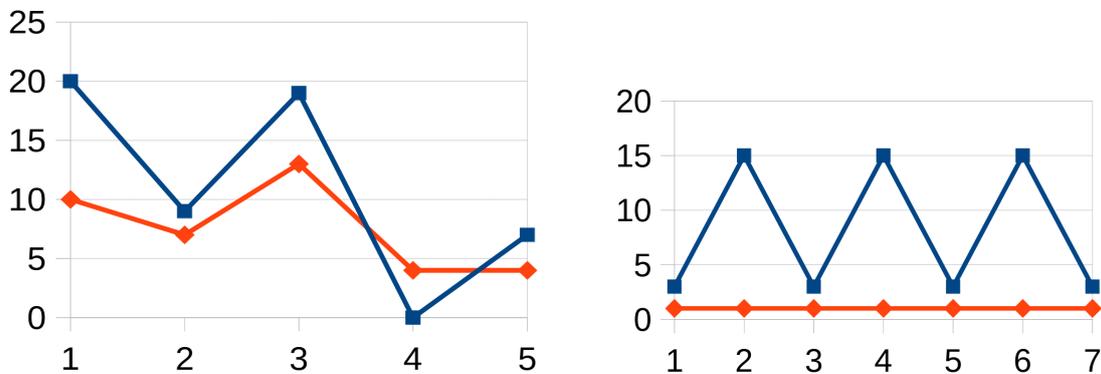


Abbildung 5: Steigung

Uns kam die Idee, dass es möglich wäre, die Graphen der Kurzzeit- und der Langzeitmessungen mithilfe der Berechnung der Steigung zuzuordnen. Wir suchten uns jeweils vier Punkte raus, berechneten die Steigung zwischen diesen und dasselbe machten wir mit den Probewerten. Wir versuchten, die Steigungen der Kurzzeit- und der Langzeitmessungen zu summieren und beide zu vergleichen. Das Problem dabei ist, dass bei unserer Theorie ein Graph, der beispielsweise aus Zacken mit den Steigungen +1 und -1 besteht, die summierte Steigung 0 hat, was einem Graphen, der konstant keine Steigung hat, gleicht. Dieser Gedanke widersprach unserer Theorie. Wir passten die Methode also an und summierten die Beträge der Steigungen. Wir hofften so die Varianz der Werte innerhalb eines Datensatzes zu quantifizieren bzw. einen Wert für die "Kantigkeit" eines Datensatzes zu bilden. Des Weiteren testeten wir, ob die Variante, die Steigung der einzelnen Channels zu vergleichen, sinnvoller ist. Dabei gilt: Je geringer die Differenz ist, desto ähnlicher ist die Steigung und somit wären die Graphen leicht zuzuordnen. Leider war auch dieser Ansatz nicht zielführend, da die Zahlen, die wir durch das Programmieren herausbekamen, eine zu große Abweichung zu den zuzuordnen-

den Werten haben und sie daher nicht aussagekräftig genug sind.

5 Analogie zu Vektorrechnung und funktionstüchtiges Verfahren

Wir haben ein Programm geschrieben, welches die Datensätze aus den jeweiligen CSV-Dateien ausliest und in Listen abspeichert. Anschließend wurden die Werte der Materialliste skaliert. Mithilfe einer Vergleichsmethode führt es unsere Rechnung mit jedem der 21 Materialdatensätze für jeweils eine Probe durch und gibt jeweils das Material aus, bei dem unsere Rechenvorschrift das größte Ergebnis liefert.

Dass uns ein systematischer Fehler unterlaufen war, wurde sofort daran deutlich, dass jede Probe von dem Programm zu dem gleichen Material zugeordnet wird.

Wir kamen schnell auf den Verdacht, dass eine fehlerhafte Normierung für das Versagen des Programms verantwortlich sei. Uns war vorher schon aufgefallen, dass unser Verfahren, der channelweisen Multiplikation dem Skalarprodukt ähnelte. Wir entschieden uns also, die Datensätze komplett als Vektoren zu interpretieren, wobei jeder Channel einer eigenen Dimension entspricht. Will man Vektoren verschiedener Längen normieren, muss man ihren Einheitsvektor bestimmen. In unserem Fall wollen wir die Einheitsvektoren der Materialkategorien bilden.

$$\vec{M}' = \frac{\vec{M}}{|\vec{M}|}$$

Wenn wir nun das Skalarprodukt vom Probevektor und Materialeinheitsvektoren bilden, bewegt sich das Ergebnis theoretisch im Bereich $[0; |\vec{P}|]$ abhängig davon, wie parallel die Vektoren zueinander stehen. Analog zu der zuvor beschriebenen Methode ordnen wir die Probe der Materialkategorie zu, bei dem das Skalarprodukt den größten Wert geliefert hat.

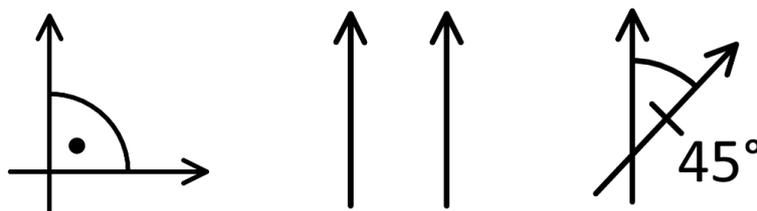


Abbildung 6: Veranschaulichung des Skalarproduktes im zweidimensionalen Raum für (i) senkrechte Vektoren (ii) parallele Vektoren und (iii) Vektoren im 45°-Winkel

6 Optimierung des Verfahrens

Um die Trefferrate unseres Verfahrens zu erhöhen, versuchten wir, die Menge an Channels zu finden, deren Betrachtung zu den besten Ergebnissen führt. Die physikalische Rechtfertigung hierfür war, dass vor allem die ersten Channel, die Gammastrahlen mit niedriger Energie detektieren, durch die Beschaffenheit des Detektors nicht so genau und zuverlässig wie die späteren Kanäle sind. Die Auflösung der Verfahren, also die Fragestellung, ob und wie viele Channel wir zusammenfassen sollten, warf sich auf, weil ein Channel bei verschiedenen Messungen nicht immer der gleichen Energie entspricht.

Den optimalen Betrachtungsbereich versuchten wir mithilfe eines Brute-force Ansatzes zu bestimmen, jedoch war die Laufzeit dieses Programms zu lang, um viele Bereiche auszuprobieren, also konnten wir nicht viele unterschiedliche Bereiche testen. Es zeichnete sich jedoch eine Tendenz ab, die die Intuition bestätigt, dass ein größerer Betrachtungsbereich vorteilhaft ist. Jedoch wurde auch klar, dass vor allem von den ersten Channels welche abgeschnitten werden mussten. Wir verblieben

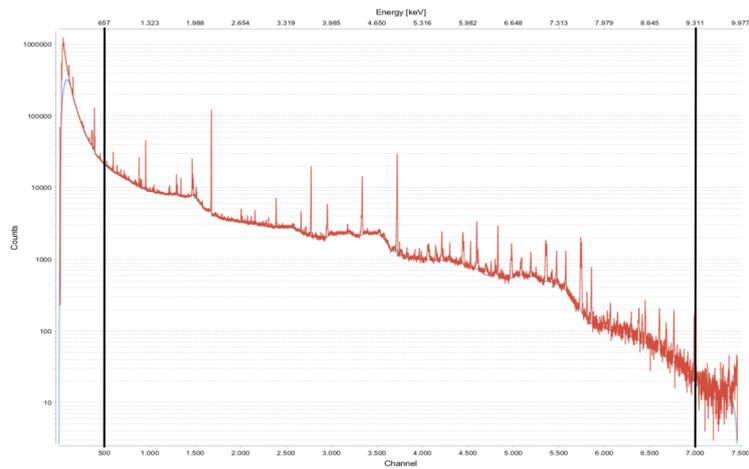


Abbildung 7: Definitionsbereich

letztendlich bei dem Bereich [500;16000]. Wir implementierten mehrere Versionen mit verschiedenen Auflösungen, bei dem Steigungs- und Vektoransatz.

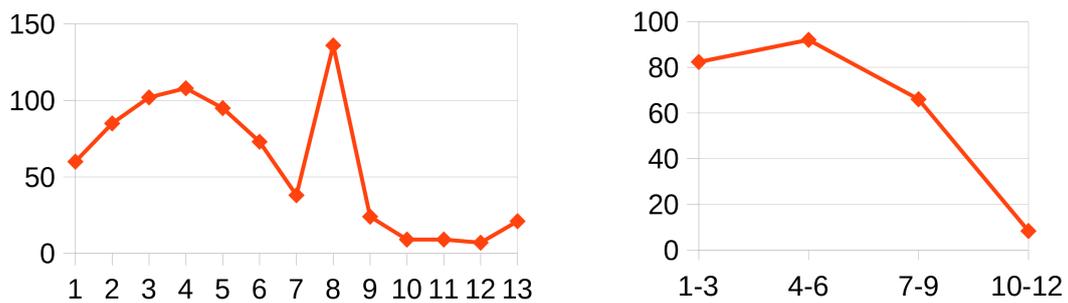


Abbildung 8: Gruppierung

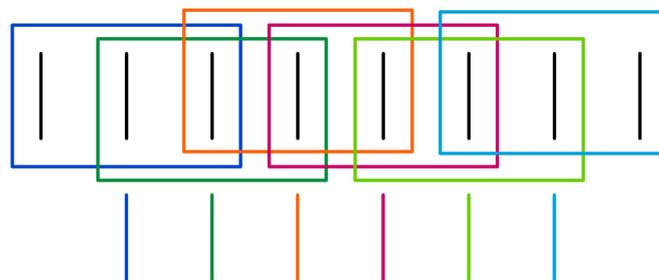


Abbildung 9: Durchschnitt

Bei dem Vektoransatz versuchten wir zu vermeiden, dass unsere Ergebnisse dadurch verfälscht werden, dass die Peaks bei zwei verschiedenen Messungen um einen Channel verschoben sind oder sich auf ein paar Channel verteilen. Um Probleme hiermit zu vermeiden, versuchten wir zum einen, mehrere Datenpunkte zu einem zusammenzufassen und zum anderen, die Datenpunkte jeweils mit dem Durchschnitt von sich selbst und den beiden Nachbarzahlen zu ersetzen.

7 Ergebnisse

7.1 Vektoransatz

Dieser Ansatz lieferte bei Berücksichtigung aller Kanäle außer den äußersten paar und ohne Voreinstellungen eine Trefferquote von knapp 85%. Als wir den Bereich optimiert hatten, erreichten wir 89,2%. Die Verringerung der Auflösung durch Zusammenfassen der Werte führte nur zu Verschlechterungen. Als wir die Auflösung gedrittelt hatten, stieg die Anzahl der Fehler ungefähr auf das Zweifache an. Als wir die Daten nach dem anderen Verfahren, die Datenpunkte jeweils mit dem Durchschnitt von sich selbst und den beiden Nachbarn zu ersetzen, modifiziert hatten, beobachteten wir ähnliche negative Effekte.

Auffällig war, dass alle Implementationen dieses Ansatzes vor allem damit Schwierigkeiten hatten, innerhalb einer Oberkategorie zu unterscheiden, vor allem die verschiedenen Kupfervarianten wurden häufig miteinander verwechselt. Mit dem besten unserer Programme passierten von unseren 27 Fehlern 8 innerhalb der Oberkategorie des Aluminiums und 10 innerhalb der von Kupfer. Dieses Muster ließ sich bei allen anderen Programmen so qualitativ beobachten.

7.2 Steigungsansatz

Wir haben letztendlich zwei steigungsbasierte Ansätze implementiert. Der Ansatz, jedem Graphen eine Kantigkeit zuzuordnen, erbrachte keine besseren Ergebnisse als rein willkürliche Zuordnung. Mit dem Ansatz, die Steigungen channelweise zu vergleichen, erreichten wir eine Trefferquote von ca. 66%.

8 Fazit

Leider konnte bei den vorliegenden Daten der Steigungsansatz nicht überzeugen. Trotz vieler verschiedener Optimierungen konnte keine überzeugende Trefferquote erreicht werden. Durch den Vektoransatz ist es uns gelungen ein Programm zu entwickeln, welches eine sehr gute Trefferquote bietet. Wenn man die Aufteilung der verschiedenen Kupfermaterialien und der verschiedenen Aluminiummaterialien aufhebt - also Kupfer und Aluminium jeweils in einer Gruppe zusammenfasst - erreicht man mit der Methode sogar eine Trefferquote von 96,4%.

Zukünftig könnte man an diesem Punkt ansetzen und Methoden entwickeln, mit denen man die Materialien innerhalb der Kupfergruppe bzw. der Aluminiumgruppe unterscheiden kann.