Chair of Design of Steels

Master's Thesis

Datafication of material mechanisms in nuclear environment

René Herbert Reich, BSc

August 2020

Declaration on oath

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich diese Arbeit selbständig verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt, und mich auch sonst keiner unerlaubten Hilfsmittel bedient habe.

Ich erkläre, dass ich die Richtlinien des Senats der Montanuniversität Leoben zu "Gute wissenschaftliche Praxis" gelesen, verstanden und befolgt habe.

Weiters erkläre ich, dass die elektronische und gedruckte Version der eingereichten wissenschaftlichen Abschlussarbeit formal und inhaltlich identisch sind.

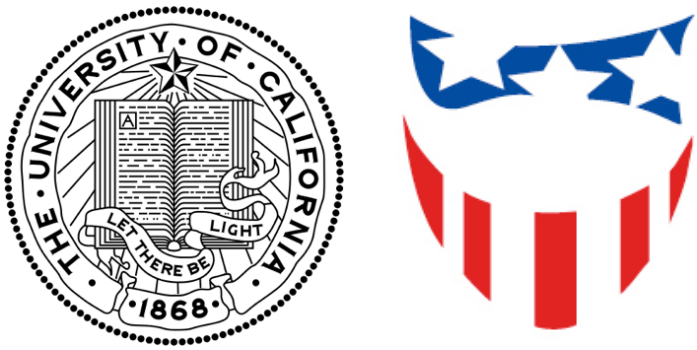
Datum 31.08.2020

René H. Reich

Unterschrift Verfasser/in
René Herbert, Reich

Acknowledgments

This thesis was carried out at the Chair of Design of Steels at the Montanuniversität Leoben, Department of Materials Science, in cooperation with the Department of Nuclear Engineering of the University of California, Berkeley, and supported by the Austrian Marshall Plan Foundation.



All my gratitude goes to the Austrian Marshall Plan Foundation and my two supervisors making this thesis possible. The four-month research stay, enabling the development of the research topic of this thesis, was supported by the grant of the Austrian Marshall Plan Foundation. Thanks to the University of California, Berkeley, its facilities and its incredible staff. Their help let me develop new skills to tackle this research problem. Thank you, Prof. DI Dr. Peter Hosemann, for welcoming me to your research group with open arms and being a great supervisor. I am incredibly thankful for Prof. DI Dr. Ronald Schnitzer, who was open to new ideas and allowed and supported me to walk the unbeaten path since I am an undergrad student.

Abstract

Material selection relies on the information of feasible materials, especially in nuclear applications, where material tests are complex and require special attention to safety. A more profound material selection decision one can make with the increasing availability of the information. Peer-reviewed scientific articles present a vital source with high quality.

This thesis studies the possibilities of information extraction and information summarization of the properties of metallic alloys. The following challenges for this purpose were identified: Fully automated information extraction by machine learning algorithms will become possible with the creation of materials science related corpus for natural language processing. Also, an automated plot digitizer would excavate enormous amounts of material data. With the increasing size of the database, its clarity decreases. So, the contained information should be summarized comprehensibly. The approach presented in this thesis is a material mechanism map. Based on the idea of Ashby maps, material mechanism maps visualize areas of material property changes under specific environmental conditions. As an example, a material mechanism map for the austenitic steel SS316, serving as nuclear fuel-cladding material, was computed. The map contains information about material hardening, recovery, irradiation embrittlement, swelling, creep, and precipitation formation depending on the irradiation dose and the homologous temperature.

Kurzfassung

Die Wahl eines Werkstoffes beruht auf den Informationen möglicher Materialien – insbesondere bei nuklearen Anwendungen, bei denen Materialprüfungen komplex sind und besondere Sicherheitsvorkehrungen erfordern. Mit zunehmender Verfügbarkeit der Informationen kann eine fundiertere Werkstoffwahl getroffen werden. Aufgrund des Begutachtungsprozesses stellen wissenschaftliche Artikel eine wichtige Informationsquelle mit hoher Qualität dar.

Diese Masterarbeit untersucht Möglichkeiten die spezifischen Kennwerte von Metalllegierungen zu extrahieren und zusammenzufassen. Diesbezüglich wurden die folgenden Herausforderungen entdeckt: Eine vollautomatische Extraktion durch Machine-Learning-Algorithmen wird erst mit der Erstellung eines werkstoffwissenschaftlichen Text-Mining-Korpus möglich. Außerdem könnte ein vollautomatischer Plot-Digitalisierer enorme Mengen an Materialdaten erheben. Mit zunehmender Größe der Datenbank nimmt ihre Übersichtlichkeit ab. Daher ist das verständliche Zusammenfassen der enthaltenen Informationen essentiell. Der in dieser Arbeit vorgeschlagene Lösungsansatz ist ein Materialmechanismen-Plot. Basierend auf der Idee der Ashby-Plots, visualisieren Materialmechanismen-Plots Bereiche bestimmter Umgebungsbedingungen, in denen sich das Materialverhalten verändert. Als Beispiel wurde ein Materialmechanismen-Plot für den austenitischen Stahl SS316, welcher als Hüllrohr für Kernbrennstäbe dient, erstellt. Der Plot enthält, in Abhängigkeit von der Strahlendosis und der homologen Temperatur, Informationen zu Verfestigung, Erholung, Versprödung, Quellen, Kriechen und Ausscheidungsbildung.

Table of contents

Declaration on oath i

Acknowledgments ii

Abstract iii

Kurzfassung iv

Table of contents v

List of figures vii

List of abbreviations xi

1. Introduction and problem statement 1

2. Fuel-cladding materials 3

3. Nuclear and elevated temperature environment 7

3.1. Irradiation swelling 8

3.2. Irradiation hardening 11

3.3. Irradiation embrittlement and fracture 14

3.4. Irradiation creep 19

3.5. Elevated temperature environment 22

4. Materials informatics 23

4.1. Machine Learning 24

4.2. Data mining 26

4.3. Text mining 27

5. Methods and computations 31

5.1. Collecting files of scientific articles 31

5.2. Machine Learning approach 33

5.3. Classification of scientific articles 34

5.4. Information extraction and consolidation 35

6. Results and discussion 38
6.1. Hardening 38
6.2. Embrittlement 41
6.3. Creep and swelling 43
6.4. Precipitation formation 44
6.5. Materials mechanism map 45
7. Conclusions 47
8. Outlook on information extraction 48
9. References 50
Appendix A: Python code collecting XML files 58
Appendix B: Python code classifying files 59
Appendix C: Python code finding entities in tables 60
Appendix D: Python code plotting the material mechanism map 62
Appendix E: Parameters of the fitted functions 64

List of figures

Fig. 1: Operating environments and considered fuel-cladding material of Gen. IV compared to Gen. III reactors, as reported in IAEA ARIS (after [2], [3]). 2

Fig. 2: Melting temperature versus neutron absorption cross-section for pure elements [6]. 4

Fig. 3: Significant alloying elements of the investigated alloys in wt.% [7], [8]. 5

Fig. 4: Swelling (in %) in pure nickel depending on the irradiation temperature at a constant neutron fluence [16]. 10

Fig. 5: Swelling as a function of neutron fluence and irradiation temperature for 20% cold-worked SS316 [20]. 10

Fig. 6: Schematic swelling behavior of SS316 and ferritic steels [21]. 11

Fig. 7: (a) The network dislocation density ρ_d at different doses and irradiation temperatures, and (b) the various contributions to yield stress change at 673K [14], [24]. 12

Fig. 8: Effect of irradiation temperature on the total dislocation density and its components in 25% cold-worked austenitic stainless steel at 7.4 dpa [19], [23]. 13

Fig. 9: Schematic stress-strain curves for fcc (a) and bcc (b) metals and the effect of irradiation with increasing dose [14]. 14

Fig. 10: The three basic fracture modes: Mode I (a), Mode II (b), and Mode III (c) [14]. 15

Fig. 11: Definition of the J-integral at the crack tip with the counterclockwise contour Γ and the traction vector T, normal to an element ds [27]. 16

Fig. 12: Schematic Charpy curves for different materials and effect of irradiation on the Charpy curve of carbon steels showing the upper-shelf energy shift (ΔUSE) and the transition temperature at 41 J shifted by ΔT_{41} 17

Fig. 13: Schematic fracture stress σ_f and yield stress σ_y curves showing the ductile-to-brittle transition temperature (DBTT) (a) and the effect of irradiation on the DBTT (b) [15], [22]. 18

Fig. 14: The very first published deformation-mechanism map by Ashby. It is the map for pure silver of grain size 32 µm and a critical strain rate εc of 10^{-8} s^{-1} [32]. 20

Fig. 15: Irradiation creep in 20% cold-worked SS316 loaded with 138 MPa at 454 °C [14]. 20

Fig. 16: Constructed deformation mechanism maps for SS316 considering just thermal creep (a) and extended by irradiation creep mechanism (b) [33]. 21

Fig. 17: Basic ML model development steps. 25

Fig. 18: Schematic workflow in supervised data mining. In the figure, the predictive model as the usage of the data mining result is just an example. 26

Fig. 19: Four different patterns (a), (b), (c), and (d) derived from the same training data set (black dots) where the x-values represent features, and the y-values represent targets [42]. 27

Fig. 20: Two examples (a) and (b) for information extraction in text mining. 28

Fig. 21: Exemplary workflow of text mining information extraction (based on [49]). 30

Fig. 22: Simple example of an XML-file with the different document parts marked. 32

Fig. 23: Natural language processing pipeline of the ChemDataExtractor [52]. 33

Fig. 24: Example of a Regular Expression and its results. 34

Fig. 25: Overview of the entities considered to describe the topic of material mechanisms in nuclear fuel-cladding applications. 35

Fig. 26: Outline of the actual 10’116 entries long spreadsheet containing the entity appearances within the corresponding table. Temp, YS, UTS, HV, UE, TE, F_tough, Embrit stand for temperature, yield strength, ultimate tensile

strength, hardness, uniform elongation, total elongation, fracture toughness, and embrittlement, respectively. Alloys* is a placeholder for the actual “SS316”, “HT9”, and “Zr4” as column names. 36

Fig. 27: Summary of the SS316 properties and their units gathered in the database. 37

Fig. 28: Dose and temperature values of the data instances gathered. 38

Fig. 29: Change of the yield strength by irradiation, showing initial hardening (a) and instant recovery (b) at elevated temperatures [24], [63]. 39

Fig. 30: Yield strength (YS) data points (in black) and the computed yield strength function σ_y(T,dpa) in blue. 40

Fig. 31: (a) Ultimate tensile strength (UTS) data points and the fitted two-variant fourth-order polynomial. (b) Comparison of the two fitted functions of YS (in orange) and UTS (in blue) with each other. 41

Fig. 32: Uniform elongation (UE) (a) and total elongation (TE) (b) data points and the fitted double polynomial planes (in gray). 42

Fig. 33: Schematic diagram of various ductility degradation processes [65]. 42

Fig. 34: Calculated dose and temperature dependence of swelling [72]. 43

Fig. 35: Density decrease data points (in black) with the fitted linear trend plane (in blue) as a 3D-plot (a) and a contour plot (b). 44

Fig. 36: Irradiation conditions for Ni_3Si (γ’) formation in SS316 [71]. 44

Fig. 37: Irradiation conditions for carbide and rod-shaped σ precipitation formations in SS316 [74]. 45

Fig. 38: Material mechanism map for SS316, giving an overview of the different mechanisms to consider at specific doses and temperatures. For example, at 0.35 T/T_M and 20 dpa (point A), one should consider swelling, hardening, the concurrence of YS and UTS, embrittlement, and irradiation creep and might neglect precipitation formations or Helium embrittlement. At 0.43 T/T_M and 55 dpa (point B), one should consider swelling, recovery and re-irradiation, embrittlement, thermal and irradiation creep, as well as the formation of carbide, σ, and γ’ precipitates. 46

Fig. 39: Overview of the working principle of the material data extractor based on the ChemDataExtractor [52]. Red fields represent the identified parts missing in materials science to achieve fully automated information extraction. 49

List of abbreviations

API	Application programming interface
ASME	American Society of Mechanical Engineers
ASTM	American Society for Testing and Materials
bcc	Body-centered-cubic
bct	Body-centered-tetragonal
Ch.	Chapter
DBTT	Ductile-brittle transition temperature
dpa	Displacement per atom
E	Energy
Eq.	Equation
fcc	Face-centered-cubic
Fig.	Figure
hcp	Hexagonal closed packed
HTML	Hypertext Markup Language
IAEA	International Atomic Energy Agency
ML	Machine Learning
NRC	Nuclear Regulatory Commission
PDF	Portable Document Format
POS	Part-Of-Speech
Regex	Regular Expression
SS316	Stainless Steel 316
T/T_M	Homologous temperature
TE	Total elongation
UE	Uniform elongation
USE	Upper shelf energy
UTS	Ultimate tensile strength
XML	Extensible Markup Language
YS	Yield strength

1. Introduction and problem statement

Material selection is a key step in the design process of a product. Its main objective is to choose the right material for a specific application with a specific set of tasks. These tasks vary widely, including bearing loads, conducting heat, resisting corrosion, and withstanding extreme temperatures. Because of the differences in the exposures' nature, materials are described by figures reflecting their different properties. Material property values represent experience from the past in the form of former applications, measurements, and material tests. They are collected for every material of interest. Thus, material selection is the step where information about the chosen material, gathered in the past, meets the predicted loads caused by the application in the future [1].

The collection of material properties is an essential aspect of material selection. The more types of materials and the more information about them are available, the better and more specific a selection one can make. Big databases are an essential medium of information collection and transportation. On the other hand, the amount of maintenance and support rises with the database size. Additionally, the essential step of information gathering is a tedious job during their setup.

The quality of the collected information ensures reliability within the selection process. Articles published in academic journals are an excellent source for material databases. The system of academic peer-reviewing ensures the reasonability of the presented data and the quality of the text explaining it. Besides the field of material selection, researchers also may profit from databases. Keeping track of the newest insights in a research field becomes increasingly harder with the developing numbers of publications.

Materials in nuclear applications have to resist extreme conditions over a long period. For example, reactor pressure vessels in nuclear power plants are designed to serve 40 to 60 years. Some of the more advanced Generation IV reactor designs are planned to serve even 80 years. In their lifetime, they have to cope with high pressures, high temperatures, corrosive environments, and radiation damage [2]–[4]. The material reliability is crucial to the safety of the plant, its workers, and its neighborhood. Further, the fuel-cladding material performance under these harsh radiation conditions is determining the time of the fuel serving in the reactor and the power plant's efficiency consequently. Thus,

material selection for nuclear applications is a step full of responsibility. On the other hand, the next nuclear reactor generation pushes to higher operating temperatures and more radiation doses, as summarized in Fig. 1. Those conditions are a challenge to the established vessel and cladding alloys. Furthermore, material properties for nuclear applications are provided in databases rarely, while irradiated material tests require additional attention to the examiner's safety [2]–[6].

Reactor type		Max. design temperature [°C]	Max. design pressure [MPa]	End-of-Life dose [dpa]	Reported cladding material
Generation III reactors					
Boiling Water Reactor	BWR	302	8.7	10	Zr-based alloy
Pressurized Water Reactor	PWR	351	17.7	100	Zr-based alloy
Generation IV reactors					
Supercritical Water-cooled Reactor	SCWR	550	27.5	67	Stainless steel
Very High-Temperature gas-cooled Reactor	VHTR	1000	10	10	TRISO
Molten Salt Reactor	MSR	700	0.5	200	TRISO, Stainless steel
Gas-cooled Fast Reactor	GFR	850	15	200	Stainless steel
Sodium-cooled Fast Reactor	SFR	550	0.3	200	Martensitic steel
Lead-cooled Fast Reactor	LFR	500	0.1	200	Martensitic steel

Fig. 1: Operating environments and considered fuel-cladding material of Gen. IV compared to Gen. III reactors, as reported in IAEA ARIS (after [2], [3]).

This master's thesis investigates the possibilities to gather information on the material properties in nuclear environments for three exemplary alloys. Fig. 1 shows that stainless steel, martensitic steel, and Zr-based alloys are the most reported fuel-cladding materials. Thus, the alloys SS316, HT-9, and Zircaloly-4 are chosen as representatives for the fuel-cladding material types, respectively. The data of the alloys should be retrieved from various scientific articles and papers in different journals. The gathered information should be merged into a publicly available database, and the different datasets should be computed to alloy specific materials property maps.

2. Fuel-cladding materials

Often the material selection is a compromise of competing properties. In nuclear applications, materials have to resist irradiation damage, high pressure, elevated temperatures, and corrosive environments.

Nuclear power plants use the energy from radioactive elements. The uranium-oxide pellets have to be safely enclosed so that they can be manipulated and do not contaminate their environment. The primary task of nuclear fuel-cladding tubes is to keep the radioactive pellets and their fission products enclosed safely, so they are not released in the reactor and its coolant. A stack of fuel pallets enclosed in a fuel-cladding tube is called “fuel rod”. Over two hundred fuel rods are tied into a fuel bundle. Boiling water or a pressurized water reactor may contain up to 256 or 289 fuel bundles, respectively. Each bundle serves between 36 to 54 months in the reactor.

When in service in the power plant, the fuel-cladding material should be as transparent as possible for neutrons to minimize neutron losses. That corresponds to a low neutron absorption cross-section of potential material. Meanwhile, the material should allow a high service temperature to enhance the thermal efficiency of the reactor. Fig. 2 is a scatter diagram of the element’s neutron absorption cross-section over the melting temperatures. Apart from carbon, the elements at the bottom of the point collection show a trend that, with increasing melting point, the neutron cross-section increases too. For fuel-cladding applications, materials out of these elements are attractive candidates, but they exhibit the disadvantageous trend of increasing cross-section with increasing melting point [6].

Further, elevated service temperatures are a challenge to the material’s integrity due to creep effects. Because of the aggressive conditions inside the tube, full of fuel pallets and fission products, and the environment of the reactor’s coolant and moderator on the outside, the tube material should be corrosion resistant on both sides. Most important for a thermal power plant, the fuel-cladding tube should conduct heat from the inside fuel to the coolant as efficiently as possible [6].

Based on the outlined considerations, martensitic steels, austenitic stainless steels, and zirconium-base alloys are primarily chosen as fuel-cladding materials. In this thesis, one alloy of each category is appointed exemplarily:

- UNS S31600, also known as EN 1.4401, X5CrNiMo17-12-2, or SAE grade 316 steel (SS316), is one of the first stainless steels developed and often used in nuclear applications.
- UNS S42100: The martensitic steel is more know under its trade name “Sandvik HT-9”.
- UNS R60804: The zirconium-base alloy, called “Zircaloy-4”, is the most frequently used fuel-cladding tube material in conventional commercial reactors.

The significant alloying elements of the enumerated materials are summarized in Fig. 3, according to the ASTM International specifications [7], [8].

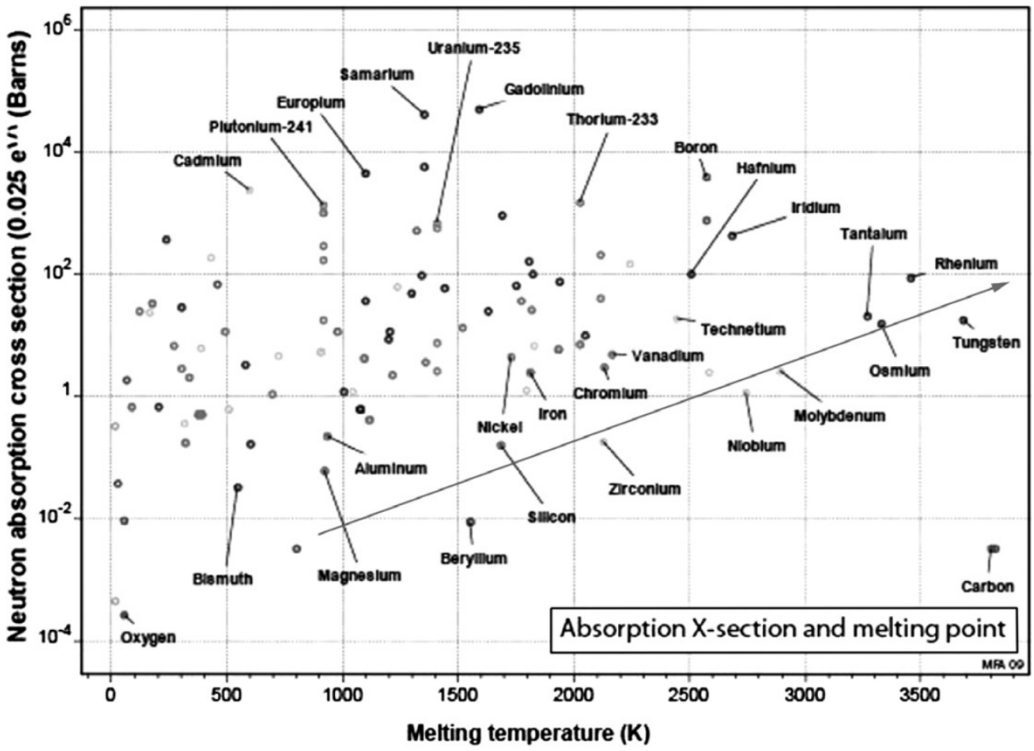


Fig. 2: Melting temperature versus neutron absorption cross-section for pure elements [6].

Depending on the temperature, alloying elements, and production’s cooling rate, steel can exhibit different lattice structures. In the absence of other alloying elements, the austenitic configuration in steel is present only between 996 K and 1766 K. One of the alloy elements that stabilizes the face-centered cubic lattice of steel down to room temperature is nickel. The high nickel content ensures that the SS316 alloy has a face-centered cubic structure at room temperature and is, therefore, austenitic steel. Because of its atomic lattice, this alloy is well-malleable. The main strengthening contribution in austenitic alloys is based on cold-work hardening.

Elements [wt. %]	UNS S31600 "SS316"	UNS S42100 "HT-9"	UNS R60804 "Zircaloy-4"
Carbon (C)	0.040 – 0.060	0.17 – 0.23	
Chromium (Cr)	17.0 – 18.0	11.0 – 12.5	0.07 – 0.13
Nickel (Ni)	13.0 – 14.0	0.30 – 0.80	
Molybdenum (Mo)	2.00 – 3.00	0.80 – 1.20	
Tungsten (W)		0.40 – 0.60	
Vanadium (V)	< 0.05	0.25 – 0.35	
Tin (Sn)			1.20 – 1.70
Iron (Fe)			0.18 – 0.24
Iron + Chromium			0.28 – 0.37

Fig. 3: Significant alloying elements of the investigated alloys in wt.% [7], [8].

HT-9 is ferritic-martensitic steel, as a result of the low nickel amount and high cooling rates in the material's heat treatment after forging and shaping. Like in a composite, the martensite is responsible for hard and brittle properties, while the ferrite is more ductile in the HT-9 structure. Until the 1970s, austenitic stainless steels were the primary fuel-cladding material. Ferritic-martensitic steels show higher thermal conductivity and lower expansion coefficients than austenites. Void swelling hinders the application of the high-swelling austenitic steels as cladding materials. On the other hand, the bcc-bct-structure and its large amount of interphases grants HT-9 excellent irradiation resistance to void swelling [9], [10], as described in Ch. 3.1.

Due to the chromium content above 10.5 wt.%, SS316 and HT-9 form a thick chromium oxide passive layer that makes them corrosion-resistant in the atmosphere. The addition of molybdenum makes steels more resistant against corrosion attacks by chlorides and reducing acids.

As shown in Fig. 2, pure zirconium has an over ten times smaller neutron absorption cross-section than iron. This excellent neutron transparency makes zirconium an interesting fuel-cladding material. Further, zirconium shows similar physical properties as iron but is more thermally stable and corrosion-resistant. Zirconium and its alloys have a hexagonal closed-packed (hcp) crystal structure. It forms a stable passive layer out of zirconium oxide on its surface, which is the reason behind the stainless property. In water, zirconium reacts to zirconium oxide under the release of hydrogen gas. This exothermic reaction is quite slow

at lower temperatures but speeds up exponentially with temperature increase. The formation of large amounts of hydrogen gas might lead to loss of contact between the cladding material and the reactor's coolant. Moreover, hydrogen is explosive in contact with the oxygen, e.g., form the air. In power plants, this scenario must be avoided [11], [12].

Alloying elements are added to increase the strength and the corrosion resistance of zirconium. Tin, chromium, and iron provide the most significant strengthening, while the neutron transparency is barely reduced. Between 200 °C and 400 °C, zirconium absorbs hydrogen, increasing the volume and embrittlement of the metal. Therefore, the amounts of nickel and iron in the Zircaloy have to be limited [11], [12].

All three described materials have in common that their corrosion and irradiation behavior are the limiting factors for fuel rods and fuel bundles used in power plants. Their swelling, corrosion, and embrittlement rates determine the service time of the fuel bundle in the reactor [13].

3. Nuclear and elevated temperature environment

Nuclear power plants make use of thermal energy from nuclear fission reactions to generate electricity. A bombardment of neutrons drives these fission reactions. A neutron can interact in different ways with the core of an atom:

- When the incoming particle transfers its complete kinetic energy to the emitting particle, it is called “elastic scattering”.
- When the emitted particle is the same as the captured one, but a loss of kinetic energy occurs during the particle transition, it is called “inelastic scattering”.
- Reactions where two neutrons are released after the neutron capture by the nucleus are called “$(n, 2n)$ reactions”. Because this reaction type produces additional neutrons, this reaction is essential to keep the chain reaction in the reactor core running.
- Nuclear reactions where a photon (γ), a proton (p), or an alpha particle (α) is emitted are called (n, γ), (n, p), or (n, α) reactions, respectively.

Based on their kinetic energy, the two most important classes of neutrons are the fast neutrons (E > 1 MeV) and the slower thermal neutrons (E = 0.025 eV). The neutron energy influences how and with which atom the neutron will interact [14]. Although thermal reactors have a mixed neutron spectrum with a wide range of neutron energies, they have to slow many neutrons down to thermal neutrons for the upkeep of the nuclear chain reaction. Fast reactors, also called “breeders”, sustain their chain reaction by fast neutrons. Nowadays, most of the nuclear power plants use thermal reactors. However, half of the prospective Generation IV reactors are fast reactors [4].

So far, just metal alloys are used for reactor pressure vessels and fuel cladding materials. Hence, the interactions of neutron radiation with metals are of particular interest. The neutron, as an energetic projectile, strikes one or several atoms in the vessel or cladding material. If the energy of the incoming neutron is high enough, the interaction can lead to the displacement of the target atom from its lattice site. The vacant site left behind is called “vacancy”. The displaced atom could interact with other atoms, find another vacancy for occupying or embed itself between regular occupied lattice sites. Atoms embedded between lattice sites are called “interstitial” atoms. The pair of a vacancy and an interstitial atom is called “Frenkel Pair”, and it is the primary cause for physical and mechanical changes of the irradiated metal [14].

The unit “displacements per atom” (dpa) was established in radiation materials science to describe the atomic damage within an alloy. It quantifies the number of displacements every atom went through statistically. The unit depends on the radiation flux ϕ, the displacement cross-section σ_D, the time the material was irradiated t, and the maximum and minimum radiation energies E_{max}, E_{min}:

$$dpa = \int_{E_{min}}^{E_{max}} \phi(E) \cdot \sigma_D(E) \cdot t \cdot dE. \qquad \text{Eq. 1}$$

The displacement cross-section σ_D is an intrinsic property of the irradiated material at specific incoming radiation energy. It indicates how much energy is transferred from the particle to the atom at the collision and how many displacements the knocked-out atom will cause. It is important to emphasize that dpa has no linear correlation with the radiation fluence because dpa includes a material characteristic. More general, the unit can be described as a ratio:

$$dpa = \frac{R}{N} \cdot t, \qquad \text{Eq. 2}$$

where R is the number of displacements per unit volume per unit time, t is the irradiation time, and N is the atom number density [14].

These ongoing displacements and the accompanied Frenkel Pairs result in physical changes of the material, such as swelling, growth, phase change, and segregation. Moreover, irradiation varies the mechanical properties such that materials perform much differently than their unirradiated equivalents. Some examples of material effects due to irradiation are swelling, hardening, embrittlement, loss of creep strength, accelerated corrosion, and intergranular cracking [14]. The following chapters describe some of these effects in more detail.

3.1. Irradiation swelling

Swelling is a dimensional instability, and it describes a change of the linear dimensions, whereas the volume of the component increases too. These dimensional changes impose a challenge on reactor engineers and designers. The volume change is caused by the formation and growth of voids and bubbles within the material. Bubbles are cavities filled with insoluble gas, which provides

internal pressure that stabilizes the cavity. Voids are empty cavities only stable under a flux of vacancies. So, voids are formed by the precipitation of irradiation-induced vacancies [15]. The volume change ΔV compared to the initial volume V can indicate swelling:

$$\frac{\Delta V}{V} = \frac{4\pi}{3} \int_0^\infty R^3 \cdot \rho_V(R)\, dR, \qquad \text{Eq. 3}$$

where $\rho_V(R)\, dR$ is the number of voids per volume with radii between R and $R + dR$. When the radius distribution is narrow, the integral can be simplified to:

$$\frac{\Delta V}{V} = \frac{4\pi}{3} \cdot \bar{R}^3 \cdot \rho_V, \qquad \text{Eq. 4}$$

where $\bar{R}$ is the mean void radius [14]. Typically, the volume change $\frac{\Delta V}{V}$ is represented in percent.

While the fast-moving interstitial atoms tend to migrate until the verge of the cascade, the slower-moving vacancies tend to form clusters nearby. So, vacancies need some mobility to cease as a void [15]. Brimhall et al. [16] showed that at low temperatures, the low defect mobility hinders void growth. High temperatures and the thermal motion of the atoms lead to vacancy annihilations, as described in Ch. 3.5. These two void destructive mechanisms result in a peak at an intermediate temperature, as shown in Fig. 4. So, the lower temperature limit for void formation is commonly set around 300 °C [17]–[19].

Garner and Gelles [20] describe a steady-state swelling rate in austenitic stainless steels around 1 %/dpa over a wide dose range between 427 °C and 650 °C. Fig. 5 displays that an onset dose is needed to start void swelling and that just the incubation dose, which lies in between the onset dose and the steady-state swelling, increases with lower irradiation temperature.

Alloy elements can affect the material's irradiation swelling behavior. In Fe-Cr-Ni austenitic alloy, swelling drops with increasing nickel content and increases with rising chromium content. Minor elements can reduce the mobility of either vacancies or interstitials by binding them with sufficient strength. Precipitates can delay void growth or initiate vacancy-interstitial annihilation. Induced stress, e.g., by cold work, reduces the transient swelling period before reaching the steady-state rate. Fig. 6 shows the difference in swelling behavior between SS316 and

ferrites schematically. Just considering swelling, the figure also indicates the advantage of the ferritic-martensitic HT-9 alloy over SS316 as fuel-cladding material [9], [14], [21].

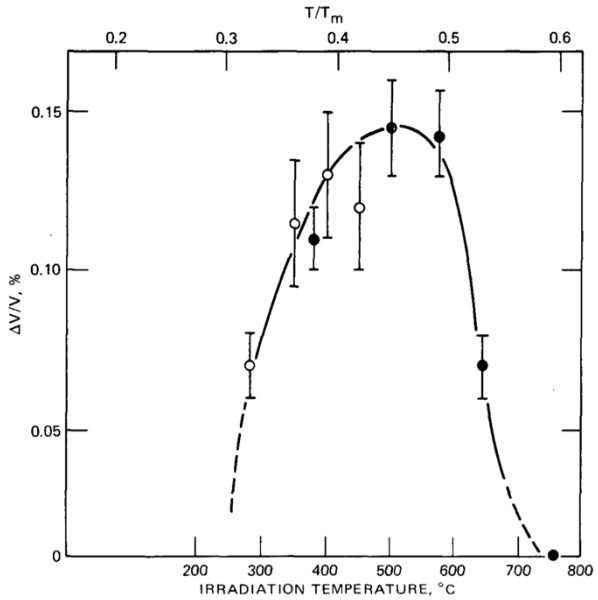


Fig. 4: Swelling (in %) in pure nickel depending on the irradiation temperature at a constant neutron fluence [16].

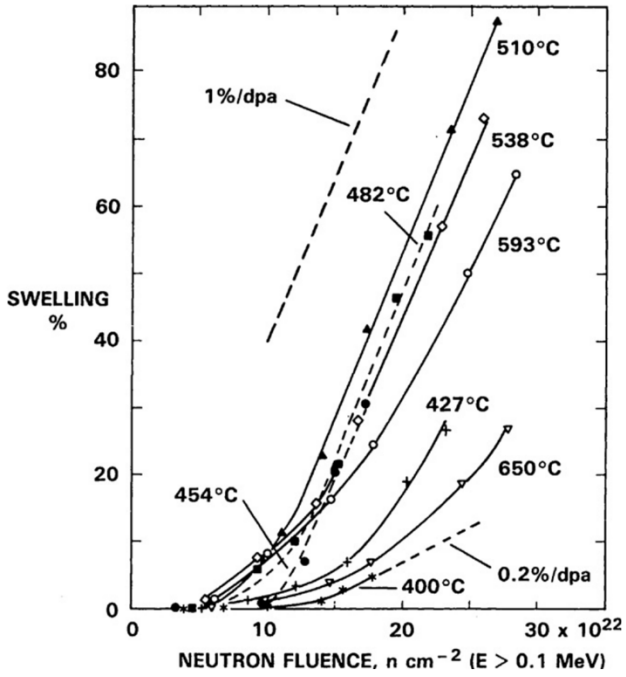


Fig. 5: Swelling as a function of neutron fluence and irradiation temperature for 20% cold-worked SS316 [20].

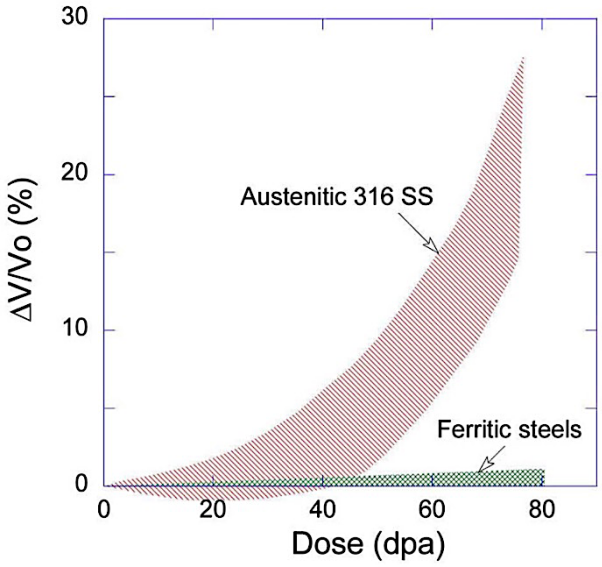


Fig. 6: Schematic swelling behavior of SS316 and ferritic steels [21].

The mechanics of voids and bubbles affect the material besides the volumetric change. Gas bubbles modify the material's physical and mechanical properties substantially. A major source for bubbles is the helium production within the alloy. Induced by thermal neutrons, boron and nickel decay emitting an alpha particle:

$$^{10}B(n,\alpha)\ ^{7}Li, \qquad \text{Eq. 5}$$

$$^{58}Ni(n,\gamma)^{59}Ni(n,\alpha)^{56}Fe. \qquad \text{Eq. 6}$$

Despite the higher dose, the material's helium swelling is lower in fast reactors than in thermal reactors. It is possible that hydrogen is generated in alloys in the same manner but by different reactions. Also, corrosion and decomposition of cooling water can produce hydrogen, which accumulates in the material's cavities [14], [22].

3.2. Irradiation hardening

The formation and mobility of dislocations determine the plasticity of a material. Thus, two types of hardening can be described [14]:

- Source hardening is the increase in unpinning stress, required to unlock a dislocation from its source and set it in motion.
- Friction hardening is the increase of flow stress, which a dislocation needs to stay in motion.

Irradiation can produce defect clusters close to Frank-Read sources. These clusters hinder the expansion of loops. Once the stress level is high enough for the loop to destroy the cluster, the dislocation can be released. So, irradiation can increase source hardening [14].

The motion of a dislocation can be hindered by obstacles in its slip plane or by other dislocations and their stress fields. Examples for obstacles are defect clusters, loops, precipitates, bubbles, or voids, and their interaction with dislocations is called “short-range stress”. The interaction of dislocations with each other or with dislocation networks is called “long-range stress” [14]. Fig. 7 (a) shows the increasing network dislocation density over the dose in austenitic steels. The figure also indicates that, below 573 K, the dislocation density increases with rising temperature. This counterintuitive behavior is a result of the different dose and temperature dependencies of the dislocation density components, which are summarized in Fig. 8 [19], [23]. The exact contributions of the different hardening mechanisms are not fully understood yet, but Lucas [24] predicted the different proportions to yield stress increase of austenitic stainless steel, as shown in Fig. 7 (b).

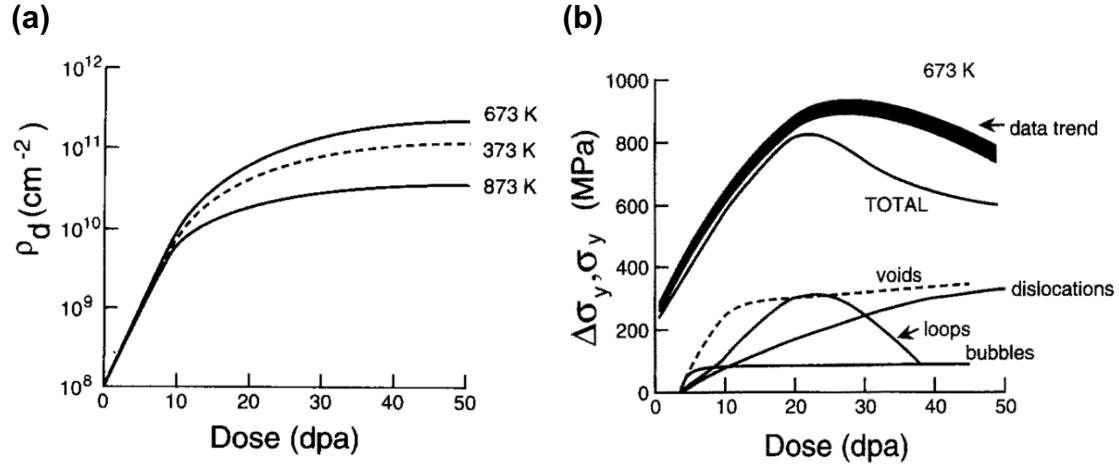


Fig. 7: (a) The network dislocation density ρ_d at different doses and irradiation temperatures, and (b) the various contributions to yield stress change at 673K [14], [24].

From Fig. 7 (a), the saturation of irradiation hardening can be derived when the creation and the destruction of obstacles equilibrate by irradiation damage, other dislocations, or temperature. If this is taken into account, the yield stress increment due to irradiation hardening can be described by an exponential formula:

$$\Delta\sigma_y = A \cdot [1 - exp(-B \cdot \phi \cdot t)]^n, \quad \text{Eq. 7}$$

where is ϕ the radiation flux, t is the irradiation time, n is an exponent to fit the model, and A and B are material-specific parameters depending on the temperature [14].

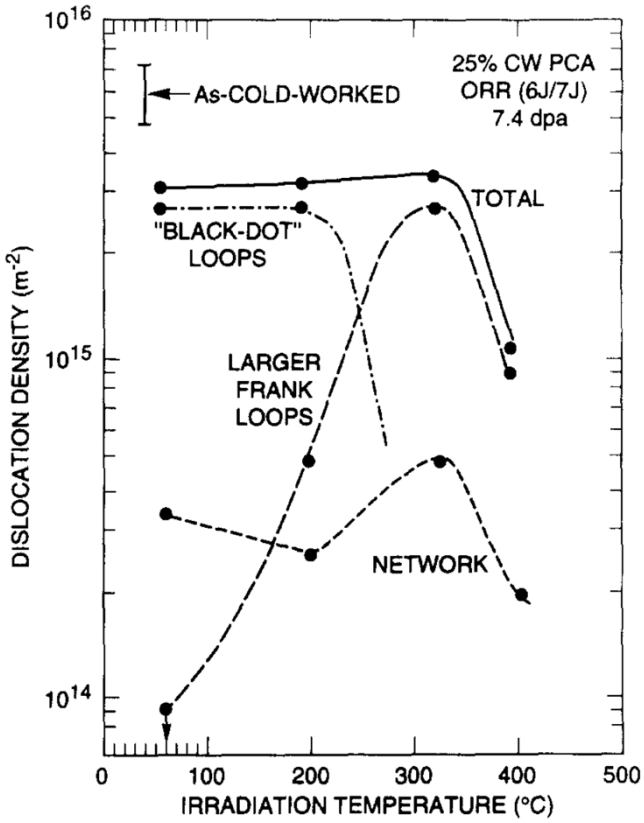


Fig. 8: Effect of irradiation temperature on the total dislocation density and its components in 25% cold-worked austenitic stainless steel at 7.4 dpa [19], [23].

Thus, irradiation hardening can be described by the increment of the metal's mechanical properties, like yield strength or ultimate tensile strength. Commonly, this material behavior is investigated in a tensile test. In a tensile test, a metal specimen is loaded with a uniaxial tensile force, while the material's elongation and the tensile force are recorded. The output can be plotted in a stress-strain curve. Fig. 9 shows two schematic stress-strain curves for fcc (a) and bcc (b) metals and the effect of irradiation with increasing dose [14].

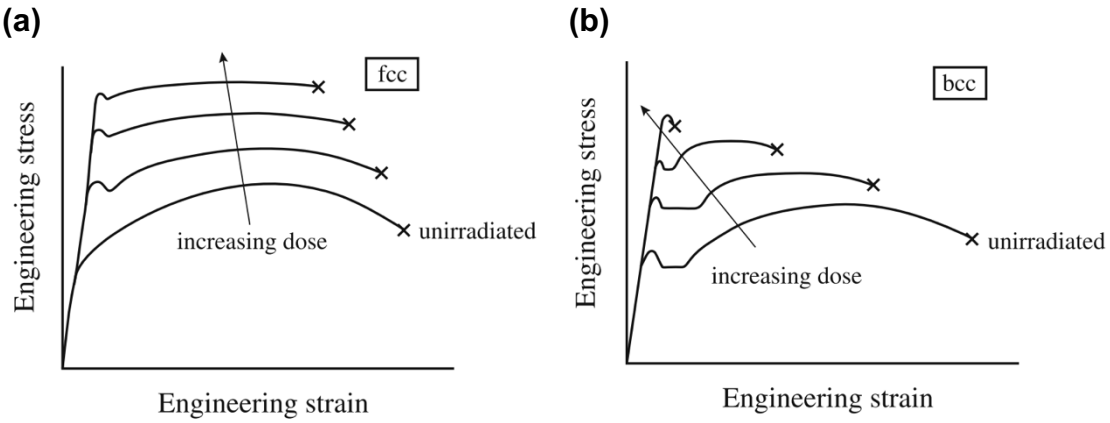


Fig. 9: Schematic stress-strain curves for fcc (a) and bcc (b) metals and the effect of irradiation with increasing dose [14].

3.3. Irradiation embrittlement and fracture

Embrittlement is the material's loss of resistance to cracking by decreasing plastic and creep deformation before fracture. Although the transition is smooth and arbitrary, two extreme types of fracture can be distinguished:

- Ductile fracture shows significant plastic deformation before and during the crack propagation.
- Brittle fracture exhibits no gross deformation, very little micro-deformation, and fast crack growth rates.

Further, fractures can be classified by the path they take through a polycrystalline material:

- Trans-granular cracks propagate through the grains.
- Inter-granular cracks propagate along the grain boundaries between the grains.

Depending on the alloy, temperature, state of stress, and loading rate, metals can show all these types of fractures [14].

Depending on the force propagating the crack, three basic modes of fracture can be defined, as depicted in Fig. 10:

(a) Mode I describes a tensile load normal to the crack plane opening the crack.
(b) In Mode II, the load shears the surfaces in-plane, normal to the crack front.
(c) In Mode III, the load shears the surfaces out of the plane, parallel to the crack front.

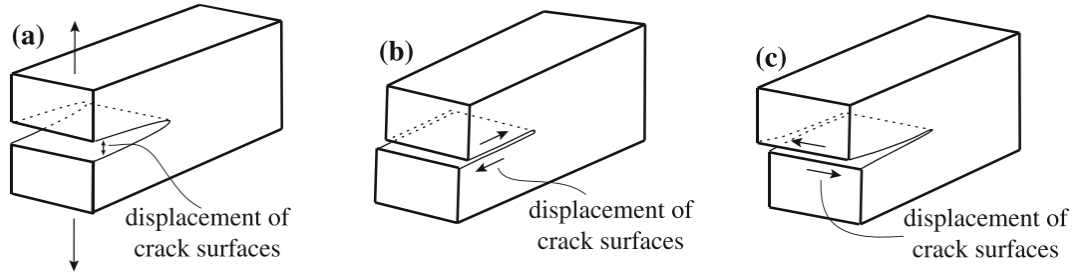


Fig. 10: The three basic fracture modes: Mode I (a), Mode II (b), and Mode III (c) [14].

In practice, Mode I is the harshest to the material, and it is, therefore, the most examined one [14].

The stress big enough to advance the fracture is called fracture stress σ_f. The crack gets propagated based on the stress field around its tip. So, to define a stress intensity factor comes handy. Notably, the critical stress intensity where the crack propagation turns into an overload fracture is essential. Based on the work of Griffith, Irwin [25] found that the critical stress intensity for a crack with a length of $2a$ in an infinite plane, which is uniform uniaxial loaded (Mode I), can be defined as:

$$K_{Ic} = \sigma_f \cdot \sqrt{\pi \cdot a}, \qquad \text{Eq. 8}$$

where K_{Ic} is also called fracture toughness. Infinite plates are hard to realize in materials test. So, usually, a geometry correction coefficient is multiplied to Eq. 8.

The fracture toughness K_{Ic} assumes that the stress states do not change at the tip of the crack. So, loads above K_{Ic} lead to brittle fracture immediately. Indeed, ceramics and steel at low temperatures show such behavior. This linear-elastic fracture mechanics is valid at high yield strength, and the component thickness is big. However, most steels behave described by elastic-plastic fracture mechanics, which also considers plastic hardening around the crack tip. To convert the geometry dependent fracture toughness to a geometry independent material characteristic for elastic-plastic behavior, Cherepanov [26] and Rice [27] suggested the so-called J-integral:

$$J = \int_{\Gamma} W\, dy - T\left(\frac{\partial u}{\partial x}\right) ds, \qquad \text{Eq. 9}$$

where Γ is a counterclockwise contour around the crack tip, W is the strain energy density, T is the traction vector normal to an element ds, and u is the displacement vector, as shown in Fig. 11.

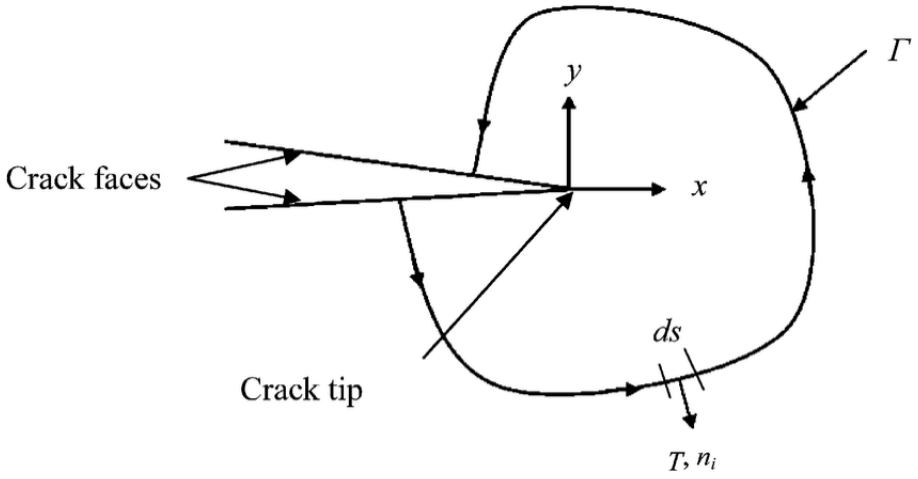


Fig. 11: Definition of the J-integral at the crack tip with the counterclockwise contour Γ and the traction vector T, normal to an element ds [27].

Physically spoken, the J-integral quantifies the difference in potential energy of two equivalent specimens having slightly different crack lengths. For elastic-plastic strain conditions, its correlation to the fracture toughness is:

$$J_{Ic} = \frac{1-\nu^2}{E} \cdot {K_{Ic}}^2, \qquad \text{Eq. 10}$$

where ν is the Poisson's ratio, and E is the elastic modulus [14].

The ASTM International standard E399 describes two sample designs to measure K_{Ic}: the 3-point loaded notched beam and the compact tension specimen. Because the fracture toughness measurements rely on the validity of linear-elastic fracture mechanics, the specimens' dimensions have to be regulated [28].

Particularly, in nuclear applications, sudden breakage of material could expose humans and the environment to danger. Thus, nuclear power plants follow the leak-before-brake concept [29]. It describes that a slowly ductile failing part

forming a leak, and warning the staff thereby, is preferable to an immediate, catastrophic breaking of the part. Accordingly, the transition from brittle to ductile behavior of the material is of particular interest. Fracture toughness tests also find the material's transition temperature, but this approach is laborious. A notched bar impact test, also called the Charpy impact test, is much simpler therefor [30]. In a Charpy impact test, a pendulum swings through a notched bar made out of the investigated material. In the impact moment of the pendulum's hammer in the bar, some kinetic energy is transferred to break the specimen. Because the weight of the pendulum is known, the height difference between the test start and after the impact quantifies the energy needed to break the material. As the geometry of the bar and the notch effect the fracture energy, they have to be standardized.

The Charpy test can be repeated with various heated or cooled bars of the same alloy. The obtained impact energies can be plotted depending on the specimens' temperatures in a Charpy impact curve, as shown in Fig. 12.

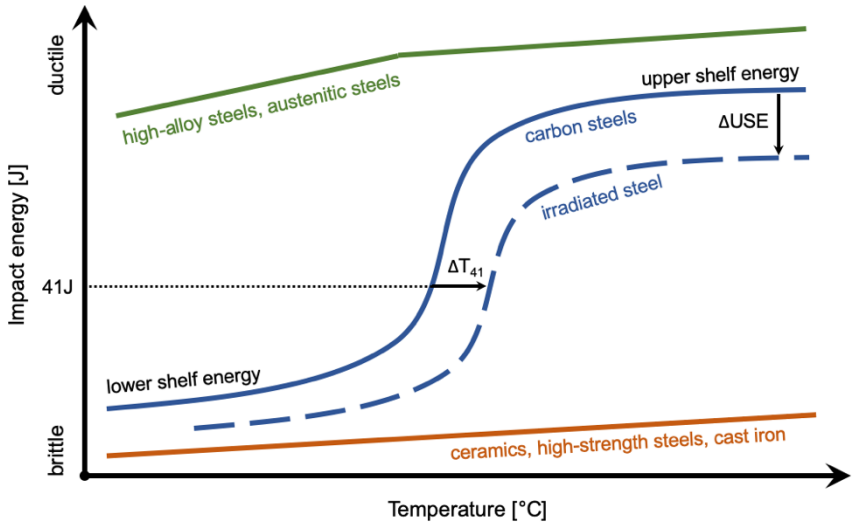


Fig. 12: Schematic Charpy curves for different materials and effect of irradiation on the Charpy curve of carbon steels showing the upper-shelf energy shift (ΔUSE) and the transition temperature at 41 J shifted by ΔT_{41}.

Fig. 12 depicts the schematic Charpy curves for various types of materials. With increasing temperature, the carbon steel Charpy curve shows the transition from brittle behavior, called lower shelf energy, to ductile behavior, called upper shelf energy. Because the brittle-ductile transition is smooth, the transition temperature has to be defined by standardized aspects. It is common to define the transition temperature at a fixed impact energy level, for example, the transition temperature at 41J of impact energy (T_{41}). The ASME BPVC III [30] and NRC 10

CFR part 50 [31] define a reference temperature of nil-ductility (RT_{NDT}), at which the specimens initiate the fracture without any plastic deformation. Based on the Charpy impact test results, the RT_{NDT} is defined by minimum impact energy and by a minimum of lateral expansion of the fractured specimens. Further, a ductile-to-brittle transition temperature (DBTT) can be defined as the intersection point of the fracture stress curve and the yield stress curve over the temperature, as shown in Fig. 13 (a) [14].

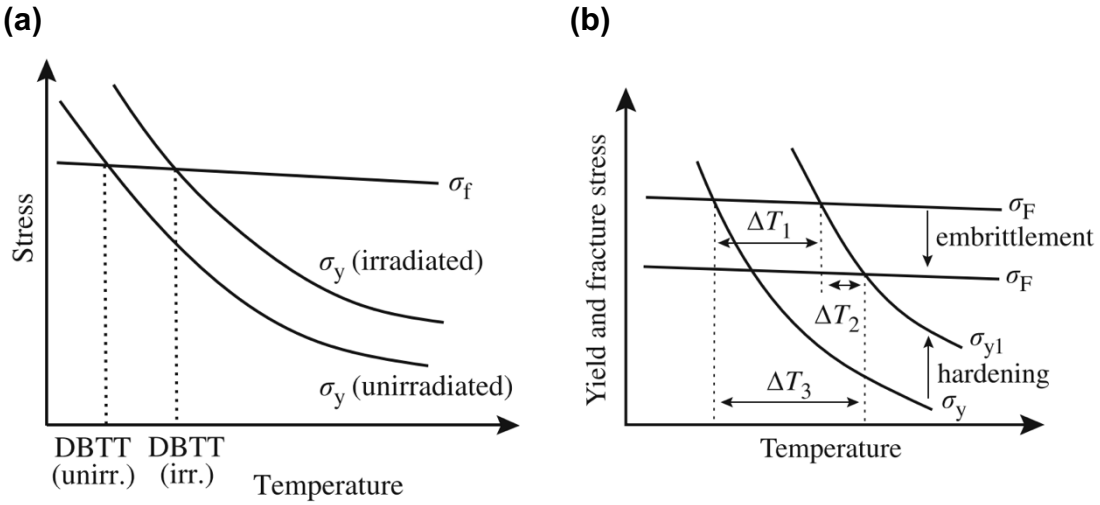


Fig. 13: Schematic fracture stress σ_f and yield stress σ_y curves showing the ductile-to-brittle transition temperature (DBTT) (a) and the effect of irradiation on the DBTT (b) [15], [22].

The neutron environment, the exposure temperature, the steel composition, and the steel microstructure affect the degree of embrittlement. The schematic Charpy curve in Fig. 12 indicates the effect of irradiation. The irradiation damage decreases the upper shelf energy by a so-called upper shelf energy shift (ΔUSE), and the transition temperature is shifted by ΔT_{41}. Irradiation can cause the formation of precipitates or the segregation of trace elements in the alloy, which both lead to a decrease of the fracture strength. Together with the yield strength increment by irradiation hardening, as described in Ch. 3.2., Fig. 13 (b) shows the transition temperature shift to higher values.

As mentioned at the end of Ch. 3.1., hydrogen can enter an alloy by corrosion, diffusion, or nuclear reactions. Besides swelling, hydrogen causes embrittlement, additionally to the direct irradiation effects. Several theories are trying to explain how hydrogen does that, reaching from the weakening of the metal-metal bonds until the formation of stress increasing bubbles. In zirconium, hydrogen forms the brittle hydride ZrH_2. Thus, hydrogen embrittlement is a crucial mechanism in ferritic steels, nickel-base, and zirconium-base alloys [14].

3.4. Irradiation creep

Creep is a time-dependent dimensional instability, like swelling, but it describes a change of linear dimension at a constant volume of the component. Without irradiation effects, creep occurs under constant load just at high temperatures in metals. At temperatures lower than 30% of the alloy's melting point ($T/T_M < 0.3$), creep can generally be neglected.

Depending on the stress and the temperature, creep is caused by two major deformation mechanisms. Ashby [32] suggested summarizing the different mechanisms in a single plot, called "deformation mechanism map". Fig. 14 shows the very first published deformation map by Ashby.

Fig. 14 illustrates that creep is the time, temperature, and stress-dependent component of plastic strain, indicated as "dislocation glide" on the map. At temperatures just above 200 °C, the field of dislocation creep emerges. Due to the high temperatures, dislocations have enough energy to climb, hindering obstacles like the ones explained in Ch. 3.2. The gliding of the so freed dislocation causes dimensional changes. Diffusional creep indicated as "diffusional flow" in Fig. 14 describes the migration of atoms in the one and vacancies in the opposite directions resulting in dimensional changes. The subfields "Coble creep" and "Nabarro-Herring creep" just indicate grain-boundary or bulk diffusion, respectively, as prevalent atom and vacancy diffusion mechanism [14], [32].

Compared to the thermal creep described so far, irradiation widens the creep domains to lower temperatures and increases the creep rate compared to thermal creep at the same temperature. As discussed in the previous chapters, irradiation influences the actions of dislocations. Thus, irradiation impacts dislocation and diffusional creep. Generally, the deformation caused by irradiation creep can be described as:

$$\varepsilon = \frac{\Delta l}{l_0} = A \cdot \left[1 - exp\left(-\frac{\phi \cdot t}{C}\right)\right] \cdot \sigma + B_0 \cdot \sigma^n \cdot \phi^m \cdot t, \qquad \text{Eq. 11}$$

where l_0 is the initial dimension, Δl is the dimension change, ϕ is the neutron flux, t is the irradiation time, σ is the stress, n and m are proportionality exponents to stress and neutron fluence respectively, and A and C are material depending

constants. The deformation ε is called tensile strain or creep strain and is described dimensionless or as percent [14].

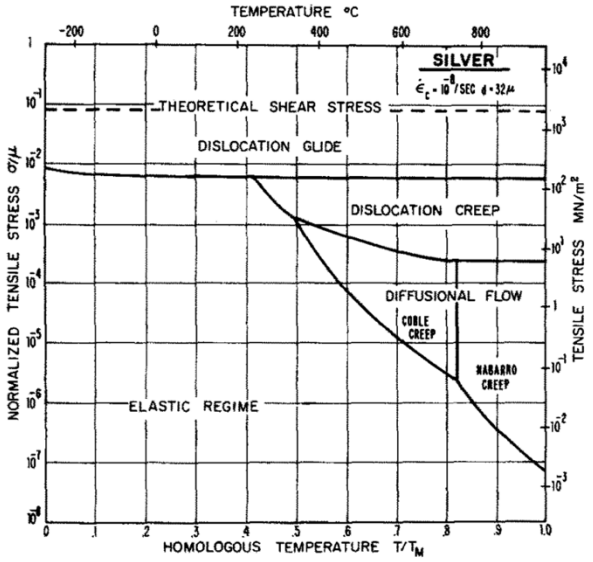


Fig. 14: The very first published deformation-mechanism map by Ashby. It is the map for pure silver of grain size 32 μm and a critical strain rate $\dot{\varepsilon}_c$ of 10^{-8} s^{-1} [32].

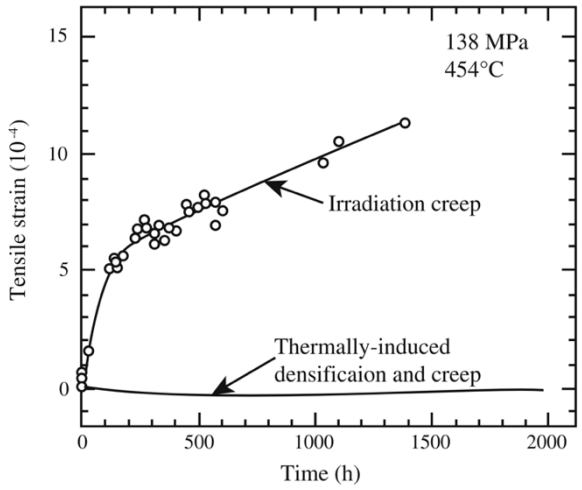


Fig. 15: Irradiation creep in 20% cold-worked SS316 loaded with 138 MPa at 454 °C [14].

Similarly to swelling, irradiation creep exhibits a transient regime, described by the first term in Eq. 11, and a steady-state regime, described by the second term. Usually, the creep strain rate depending on time ε' is of interest, but irradiation creep is often reported as an effective strain rate per dpa $\dot{\varepsilon}_{eff}$ divided unit of effective stress σ_{eff}:

$$\frac{\dot{\varepsilon}_{eff}}{\sigma_{eff}} = B_0, \qquad \text{Eq. 12}$$

where B_0 is called “creep compliance”. If void swelling is considered too, Eq. 12 can be extended to the following empirical equation:

$$\frac{\dot{\varepsilon}_{eff}}{\sigma_{eff}} = B_0 + D \cdot \dot{S}, \qquad \text{Eq. 13}$$

where $\dot{S}$ is the instantaneous volumetric swelling rate per dpa, and D is the creep-swelling coupling coefficient [14]. The irradiation creep mechanism can be added to Ashby’s deformation mechanism map, as shown for SS316 in Fig. 16 [33].

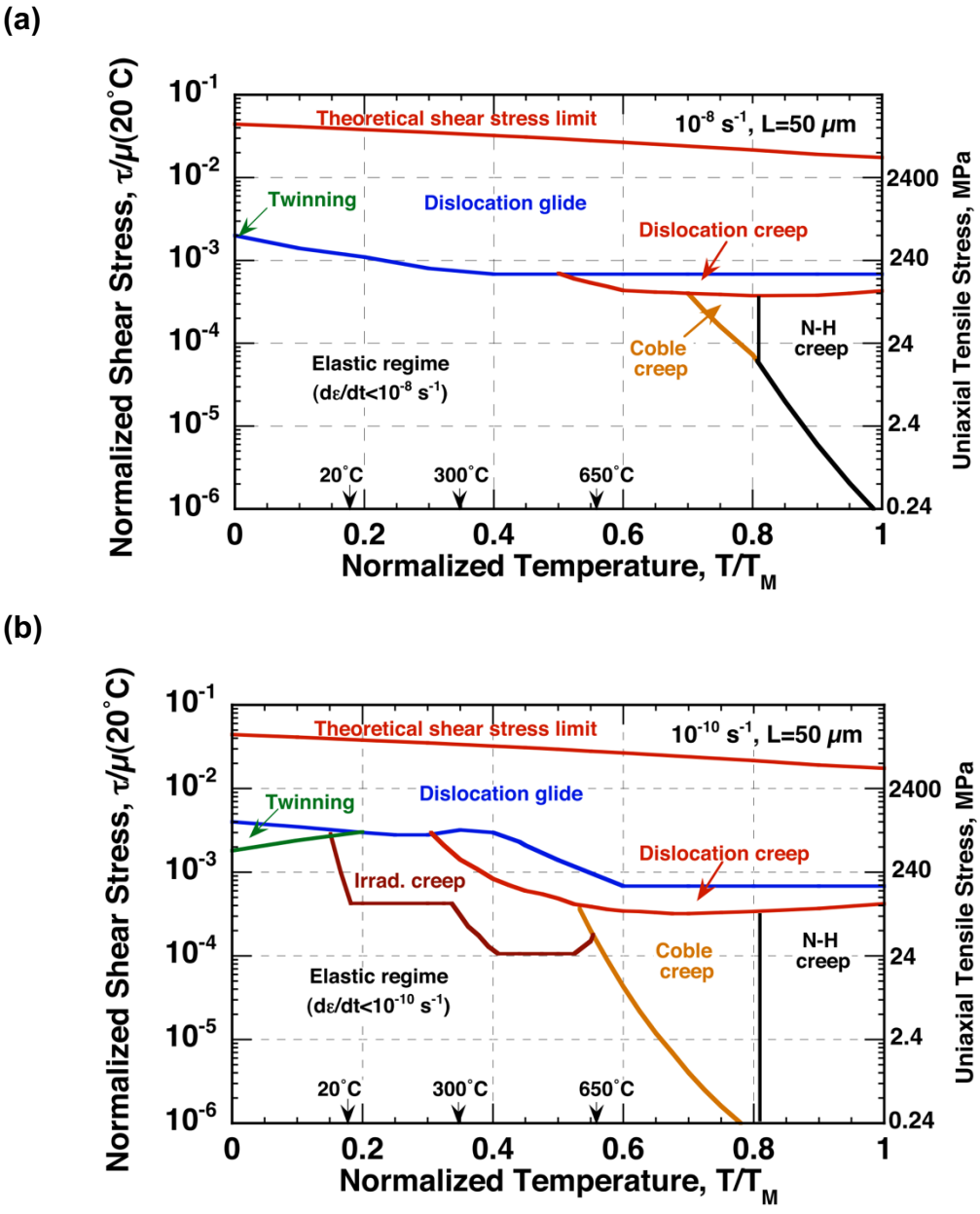


Fig. 16: Constructed deformation mechanism maps for SS316 considering just thermal creep (a) and extended by irradiation creep mechanism (b) [33].

3.5. Elevated temperature environment

The principal part of the strength in austenitic alloys, like SS316, and in zirconium-based alloys, like Zircaloy-4, is caused by what is called work-hardening. Plastic shaping of the material, e.g., bending or rolling, initiates new defects and dislocations in the crystal. In the case of a sufficient number of dislocations generated, they hinder each other mutually in their movement so that a strength improvement occurs.

With increasing temperature, the atom vibrations and movement in the steel's crystal lattice increase too. Above temperatures of 30% of the melting point ($T/T_M > 0.3$), there is enough energy to heal point defects, like interstitials and vacancies, and anneal dislocations. These processes in the material are resumed as "recovery", which results in a decrease of hardness and yield strength and in an increase of ductility. The enhanced atomic mobility causes grain growth, especially if the materials are exposed to even higher temperatures over a long period. According to the Hall-Petch equation [14], an increasing grain size lowers the material strength.

Ferritic-martensitic alloys, like HT-9, are produced by quenching, starting from temperatures where the steel is still austenitic. Due to the rapid cooling rates, the austenite forms the metastable martensite (bct) structure. The alloy's primary source of strength is the brittle martensite grains. These alloys also exhibit recovery, but more crucial are temperatures above the austenitic transformation temperature (approx. 700 °C). Then, the ferritic-martensitic structure recrystallizes, causing the material's loss of hardness and strength.

In alloys with high carbon content, chromium carbide precipitates above temperatures of 500 °C. These carbides impoverish the surrounding areas that then shall have no passivation corrosion protection due to the local lack of chromium. Moreover, the formed carbide precipitations lead to undesired embrittlement of the material.

4. Materials informatics

Materials informatics is a research field combining practices of informatics, data science, materials science, and engineering to improve the discovery, development, selection, production, use, and recycling of materials. It analyses complex, multiscale information about material production, physical or chemical properties, measured, and collected before. Materials informatics can also be used to generate and manage material data, besides its utilization. The materials informatics' objective is to compute statistically robust, physically, and chemically meaningful models to enhance one or more episodes in the material's life cycle [34]–[37].

While developing a new consumer product takes two to five years, implementing new materials may take 15 to 20 years from invention till commercial launch [38]. Modern fabrication methods, like Additive Manufacturing, enable new degrees of freedom and rocketing speed of customization. These manufacturing opportunities equal challenges for traditional materials science approaches to determine material processes and performance limits [37], [39]. In 2011, the US government introduced a multi-agency initiative, called "Materials Genome Initiative" (MGI). According to the 2014 MGI Strategic Plan, three of the four MGI's key challenges are within the materials informatics field [39].

One of the earliest examples of materials informatics is a thermodynamic database. Based on the elements of the periodic table, data about the thermodynamic contribution of every element were collected. Such databases were the fundament of thermochemical computations to map phase stabilities in binary and ternary alloys. As a result, computationally derived phase diagrams are well established in material development nowadays. Other examples of public databases in materials science are crystallographic databases, like the Inorganic Chemistry Structural Database (ICSD), or the Cambridge Structural Database (CSD) [34], [36]. It is striking that mainly digital databases with chemical information only on an atomic level were developed. Especially on the level of industrial materials and alloys, databases about corrosion behavior, processing properties, physical and mechanical properties, like yield strength or uniform elongation, are missing or just emerging [36], [37].

Materials informatics can fulfill several tasks in different steps of the material's life cycle. Rickman et al. [36] enumerate in detail a couple of application examples

reaching from material discovery, ab initio calculations, multimodal imaging, phase characterization, material optimization, till application of density-functional theory.

New material development is time-consuming, risky, and expansive. Not every new application needs a newly developed material. For example, the superalloy “Inconel 625” was initially developed as a structural material for supercritical steam power plants. It is famous for its high strength and corrosion resistance in highly acidic environments. However, the alloy found use in battery contacts of the Tesla Model S, due to its good stress response at extreme temperatures resulting from resistive heating of the contact at rapid acceleration [37], [40]. With more and more alloys invented, the range of available materials becomes unimaginable. So, selecting an optimal solution for a specific application turns tedious. Mulholland and Paradiso [37] explain how materials informatics can be used for material selection. Based on the application’s requirements and loads, machine learning tools can find an optimal candidate out of a material database.

The following three subchapters describe general concepts and methods originating outside the field of materials informatics but become essential therein.

4.1. Machine Learning

In standard programming, one defines operating instructions step-by-step, resulting in an algorithm solving a specific problem. In machine learning (ML), the computer uses a set of statistical models to compute its own algorithm based on the data provided. Machine learning is a branch of artificial intelligence. It “learns” from situations from the past to estimate a model, which can fulfill different types of tasks, like classification, regression, or clustering [41].

In general, every machine learning process undergoes the following three iterative steps [41], as summarized in Fig. 17:

(1) **Representation**: First, the task’s object or instance needs to be described so that the computer can understand it. The instance’s attribute together with its value is called “feature”, e.g. “pixels = 1200”, “diameter = 15.8”, or “yield strength = 230”. In the representation step, one has to choose which and how many features to include. Further, it has to be decided which mathematical model fits best for the task. Then the ML model can be

computed based on the input features. The input data teaching the ML model is called “training data”.

(2) **Evaluation**: In this step, one has to define what criterion distinguishes a good from a bad ML model.

(3) **Optimization**: Which changes will improve the model’s criterion? This question might be answered by changing the number of features, the mathematical model, or its parameters.

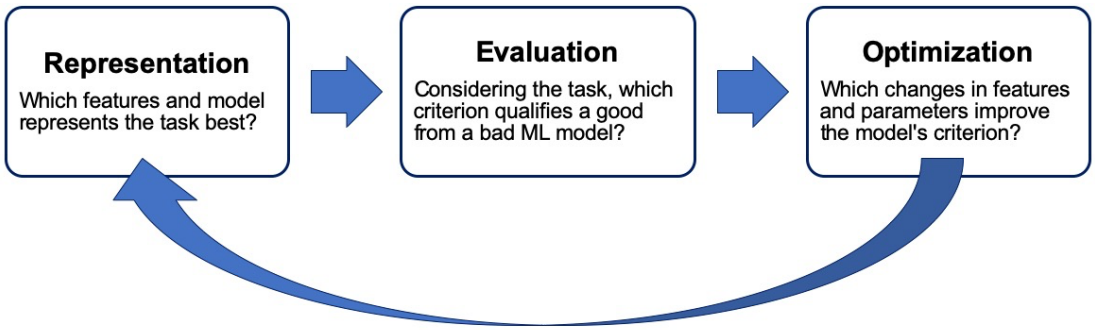


Fig. 17: Basic ML model development steps.

With the increasing number of tasks, many different systems of machine learning developed. So, it is useful to classify them in broad categories by different aspects. Just considering whether the ML model sees human supervision during its training, there are four major categories [41], [42]:

- **Supervised learning**: In supervised learning, the input data already includes the desired output, called “label” or “target”. A typical example of that is an email spam filter where the user tells the program which emails are spam or not, and the filter classifies the new incoming emails then.
- **Unsupervised learning**: As the exact opposite of supervised learning, the input data is missing information about the output totally. This approach might be used to detect clusters of voids in material or to detect anomalies.
- **Semi-supervised learning**: Because labeling is often time-consuming and costly, semi-supervised learning lays between the two categories above, where just some instances are labeled.
- **Reinforcement learning**: The ML model perceives the environment to select and perform an action then. Based on the action, the ML model gets rewards or penalties in return, and it tries to find the best strategy to maximize the rewards. Reinforcement learning is often used in robots to learn how to walk.

Jin et al. [43] used machine learning methods to predict the onset dose for void swelling (shown in Fig. 5), based on the alloy's chemical composition. Zhang et al. [44] were able to apply machine learning to predict solid solubility based on the Hume-Rothery rules with decent precision. His machine learning algorithm derived similar rules like Hume-Rothery, but slightly different parameters. Jin's and Zhang's works are a vivid example of structure-property linkage in materials informatics.

4.2. Data mining

Data mining is a branch of machine learning, but it has several vague definitions. Some use it as a synonym for knowledge discovery from data (KDD); others see it as an essential step in the process of knowledge discovery. In general, data mining uses mathematical models to find patterns in large datasets. Then, these patterns can be used to detect clusters, associations, or anomalies. They also can be used to make predictions for new observations. Based on the Cross-Industry Standard Process for Data Mining (CRISP-DM), the data mining workflow can be summarized as depicted in Fig. 18 [34], [45]–[47].

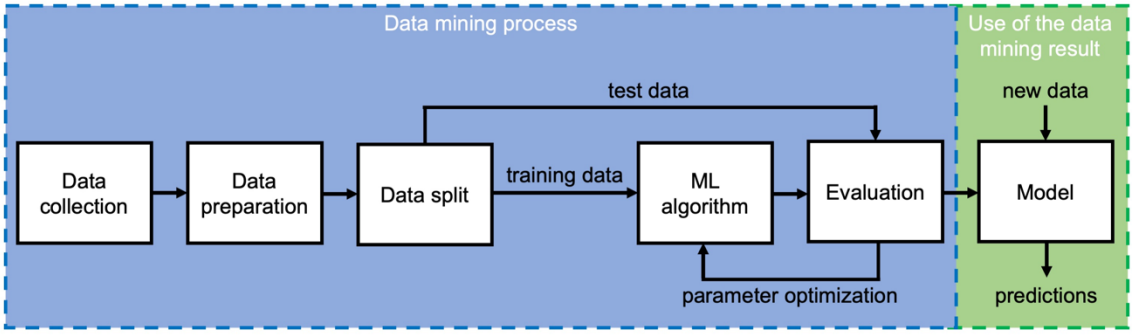


Fig. 18: Schematic workflow in supervised data mining. In the figure, the predictive model as the usage of the data mining result is just an example.

Depending on the desired output, meaningful data in sufficient amounts has to be collected first. Often, data is collected in a deductive database consisting of one or several tables. Typically, each table exhibits a set of attributes as columns and a broad set of observations or measurements as rows. Then, the dataset has to be prepared to clear noisy and inconsistent data and to replace missing values. Especially in supervised learning, a subset of the observations has to be reserved for the later evaluation. This split is done randomly to assess the algorithm's performance independently later. The ML method is then trained by the training

data only. One can choose from numerous types of ML methods, all having pros and cons in precision, computational speed, and data amount handled. Fig. 19 underlines the importance of choosing the appropriate method by showing how different patterns can be derived from the same training dataset conceptionally [45], [47].

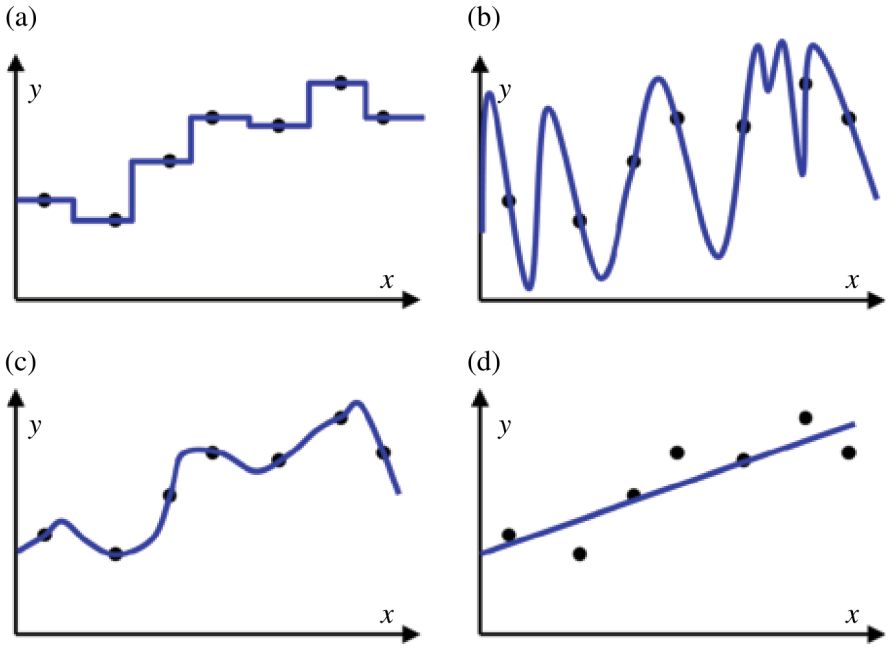


Fig. 19: Four different patterns (a), (b), (c), and (d) derived from the same training data set (black dots) where the x-values represent features, and the y-values represent targets [42].

Depending on the ML task, there is a wide range of metrics to evaluate the performance of a trained ML algorithm. In supervised learning, the algorithm can be trailed with the features contained in the test data. Because it is unfamiliar with the test dataset, the algorithm predicts the target values, which one can compare with the known correct targets afterward. This insight can be used to optimize the algorithm's parameters in an iterative process until the desired accuracy is reached [45].

4.3. Text mining

Like data mining, text mining is also a branch of machine learning. It has a variety of definitions. Similarly, text mining can be seen as a process within KDD or as the extraction of information from written text. Further, a text itself can be represented by data, and methods of data mining might be applied to find text patterns. This application is often called "text data mining" [48].

This thesis focus on information extraction from the written text published in academic articles and papers. Therefore, it is important to understand the two different forms of how information is provided. Structured data describes the regular and predictable organized form of information, like tables or a set of lists. So, tables link pieces of information together, called “relation”. Formulated sentences adding up to text is unstructured data, like this master thesis containing much information spread over several paragraphs, chapters, and pages. Unstructured data cannot be processed digitally, however. Thus, the main goal of information extraction by text mining is to translate unstructured in machine-readable structured data, as summarized in the visual examples in Fig. 20 [49].

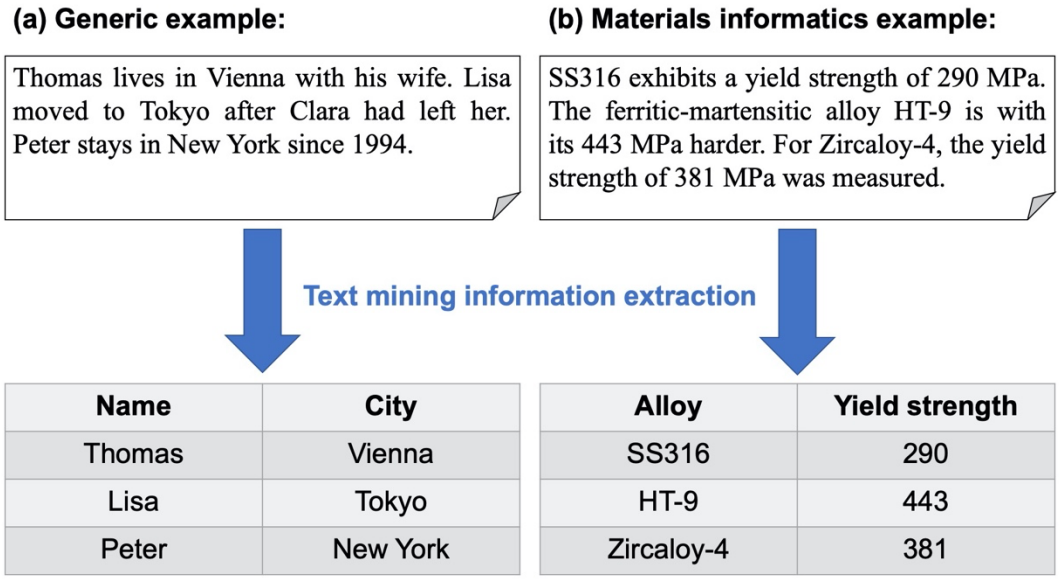


Fig. 20: Two examples (a) and (b) for information extraction in text mining.

In the generic example of Fig. 20 (a), one is interested in the cities the people live in. So, the information of the people’s names and their localization needs to be retrieved from the text. Along with this example, some essential aspects of the workflow of text mining information extraction can be outlined, as depicted in Fig. 21.

The text has to be broken into its individual sentences to make unstructured text processable for computers. Every sentence becomes an entry in a list. Then the sentences are split into the words they consist of, also called “tokens”. Now, the text is represented as a list of sentences, where every list element is a list of words itself [49].

Every word can be classified in word categories, like nouns, verbs, adjectives, adverbs, etc. This classification step is called “part-of-speech (POS) tagging” and is often done with the help of freely available online libraries, dictionaries, and programs. For a computer, this step is not trivial, as the word category is determined by context many times. Words like “current”, “list”, or “break” have different meanings and thus can be used as nouns, verbs, or adjectives. Because POS-tagging is a classification task, machine learning methods are applied to identify the correct word category [49].

Depending on the desired output, the information of interest, called “entities”, has to be found. Phrases like “New York” or “yield strength” consist out of two or more words but describe one entity. Their detection can be secured by libraries and dictionaries too [49]. In the example of Fig. 21, all the city names are detected by matching with an online library, and all the people’s names are determined as the other output class, e.g., simply by their capital letter.

Finally, the relations described within the sentences have to be recognized as well to distinguish related entities from unrelated ones [49]. Fig. 21 shows that Clara is identified as an entity precisely because it is a name. However, her name is not incorporated in the database correctly because of the missing relation to a city.

The example of Fig. 21 also clarifies that the first steps, namely sentence segmentation, tokenization, and POS-tagging, are quite standardized and universal in text mining. The logic of entity and relation recognition have to be customized to the information extraction task.

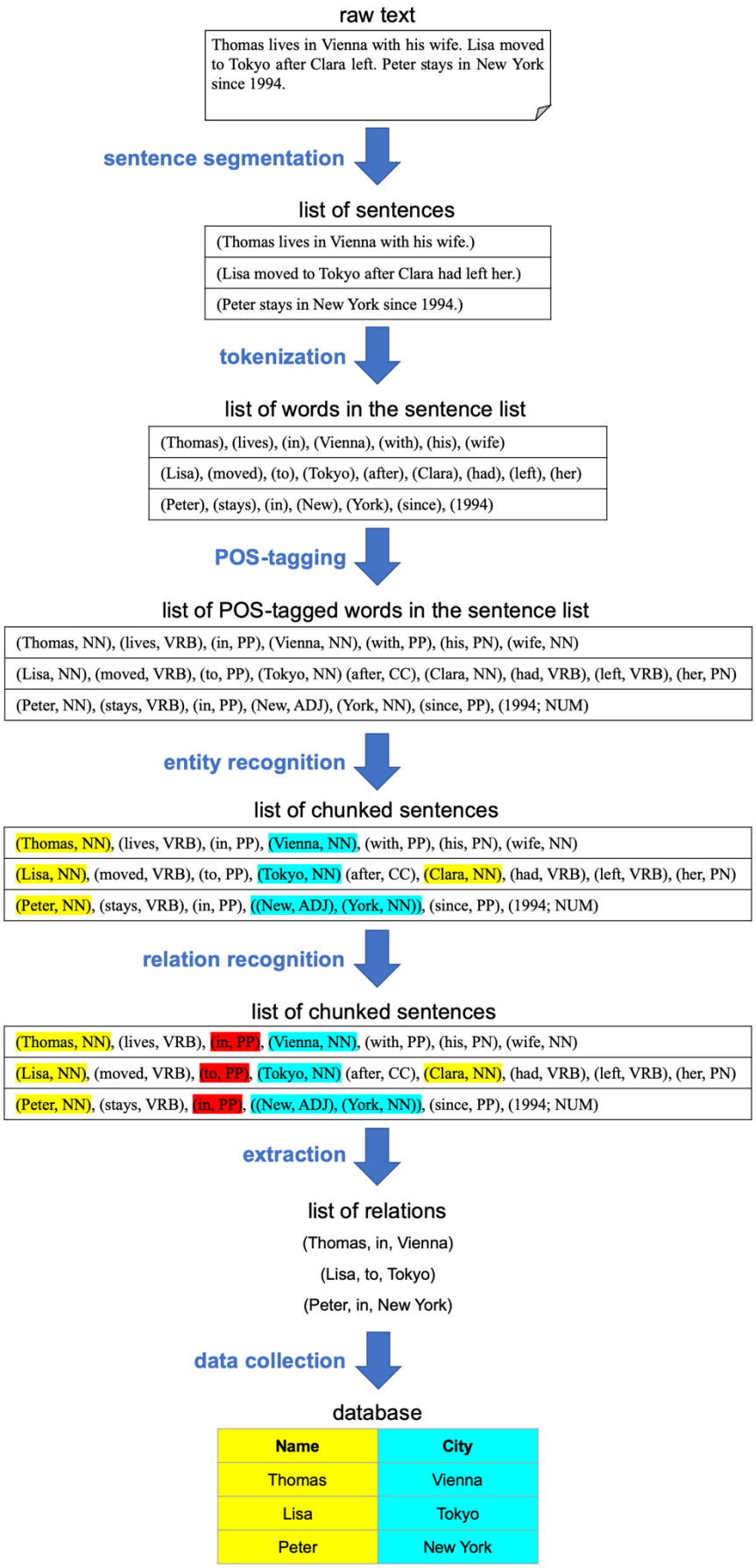


Fig. 21: Exemplary workflow of text mining information extraction (based on [49]).

5. Methods and computations

The following chapters describe the methods considered and performed for information extraction from scientific articles. The algorithms were programmed in Python. The code and the computed spreadsheets are provided in Appendix A until Appendix D, and they are publicly available at [50].

5.1. Collecting files of scientific articles

In many materials science research projects, material properties are determined. Then, the results are published in the relevant journals. Nowadays, their articles are accessed online. The reader retrieves a digital version of them, mostly in a Portable Document Format (PDF). With the evolution of text and data mining, the major publishers start to offer an Application Developer Interface (API), so researches can access the papers by programming code. Thereby, publishers turned into a very interesting source for gathering material properties.

Scientific papers can have four parts containing information: the full-text field, tables, figures, and the paper's metadata, thereunder the title, the abstract, the list of authors, or the publishing date. In 1993, Adobe Systems Incorporated developed the PDF with the goal "to enable users to exchange and view electronic documents easily and reliably, independent of the environment in which they were created or the environment in which they are viewed or printed" [51]. So, the PDF can be described as an electronic paper. It was never meant to transport or give access to datasets, but to represent them in a nice and readable way. Papers published before the internet are digitalized as PDFs mostly. Thus, the majority of scientific output from the last 40 years is distributed online in PDFs. On the other hand, HTML and XML-files provide data hierarchically structured, as shown in Fig. 22. Many papers of the last 15 years are available as such files where the paper's parts, like the title, tables, and text body, are marked as such by a specific syntax. These file formats do not contain figures and might provide an online link to the figures instead.

```
<?xml version="1.0" encoding="UTF-8"?><document>
  <metadata>
    <title>Datafication of material mechanisms in nuclear
    enviornment</title>
    <author>Rene H. Reich</author>
    <type>Master's thesis</type>
    <abstract>This thesis focuses on...</abstract>
  </metadata>
  <full-text>
    <text-body>Material selection is a key step in the
    design process of a product. Its main objective...

</text-body></full-text></document>
```

Fig. 22: Simple example of an XML-file with the different document parts marked.

For this thesis, the publishers Elsevier and Springer provided an API with full-text access to the articles available through the UC Berkeley library. Based on the query “Irradiation material test”, 13’641 and 12’571 full-text XML-files were retrieved from Elsevier and Springer, respectively. Because Taylor & Francis Online does not offer an API yet, 17’421 articles of their journals “Nuclear Technology”, “International Materials Reviews”, “Nuclear Science and Engineering”, and “Journal of Nuclear Science and Technology” were transferred after request. All articles were PDF-files, and their metadata was summarized in XML-files additionally. A minority of the XML-files even contained the paper’s full-text. As four articles were found at Elsevier as well as at Springer, a total of 43’629 XML-files, mostly containing full-texts of scientific papers, were gathered from the three mentioned publishers.

In their work, Swain and Cole [52] have explained how to set up a program to extract chemical information automatically from the scientific literature. Their ChemDataExtractor yields chemical names, alphanumeric labels, spectroscopy attributes, and chemical properties out of HTML, XML, and PDF files. While processing the HTML and XML-files is working well and achieves stunning results, Swain reports limited support for PDFs. If at all, the extractor can process PDFs of a few subject relevant journals, and even then, the process seems to be unstable. Court and Cole [53] auto-generated a database of Curie and Néel temperatures of magnetic materials. Although their program is based on the ChemDataExtractor, PDF processing was disregarded totally.

While the conversion of an XML to a clearly readable PDF is doable, the opposite way of PDFs into HTMLs or XMLs with marked document parts imposes big challenges to the informatics field. Because publishers use different layouts for their titles, graphs, figures, and tables as part of their corporate design, breaking up the PDF-document and marking the different parts correctly is the key issue.

Constantin et al. [54] tried to tackle this challenge with his converter “PDFX”. Also, PubGenius Inc. offers the payable program “Typeset” as a solution [55]. Both programs were tested by converting the PDF of [56], containing mathematic formulas, tables, and figures. Although Typeset was able to label metadata, headlines, text bodies, and figures correctly, both programs could not detect tables and convert formulas. Thus, just the XML-files were considered in the following steps for information extraction.

5.2. Machine Learning approach

Fig. 23 summarizes the natural language processing pipeline of the ChemDataExtractor [52] used to extract information out of the articles. For the steps of sentence splitting and tokenization, reliable algorithms are available. POS-tagging and entity recognition are crucial and challenging steps. POS-taggers are ML algorithms trained on already manually tagged texts. The vast majority of POS-taggers have been trained on newspaper articles, like the Wall Street Journal (WSJ) corpus. Additional to the available WSJ corpus, the tagger of the ChemDataExtractor was trained on features of some chemistry articles and on the GENIA corpus, which contains tags of 2’000 Medline abstracts. This improves the performance of the tagger, especially in the biomedical and chemistry domain. Similarly, the entity recognizer is also a machine learning algorithm trained on the CHEMDNER corpus consisting of 10’000 PubMed abstracts with 84’355 manually annotated chemical entity mentions.

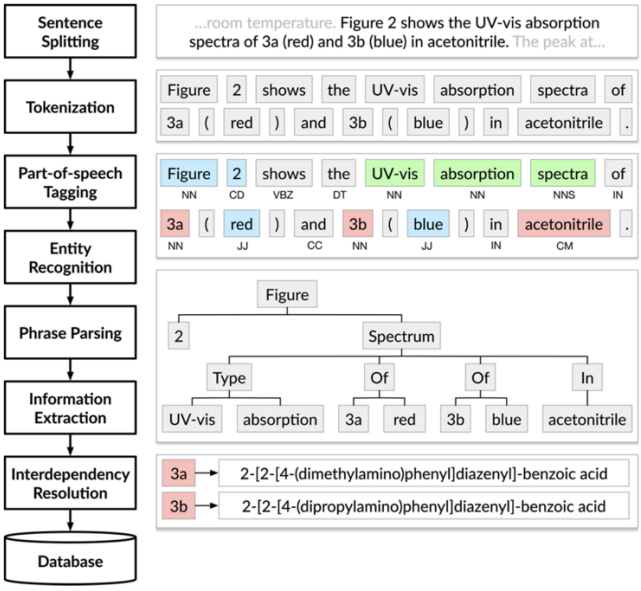


Fig. 23: Natural language processing pipeline of the ChemDataExtractor [52].

The available algorithms, especially for entity recognition, achieved unsatisfying scores for materials science and mechanical property entities. These results made clear that an ML text mining approach would be beyond the scope of this thesis, as materials science, particularly in the metal alloys field, is missing such essential corpora for this task. Thus, text mining on the full-text bodies of the papers became even more challenging.

5.3. Classification of scientific articles

The appearances of the three alloy names “SS316”, “HT9, and “Zircaloy-4” were searched in the full-texts of the 43’629 files to reduce processing times in the following steps. Therefore, the alloys, their abbreviations, and synonyms were described by Regular Expressions. Regular Expressions (Regex) enable defining a pattern out of a sequence of characters and syntactical rules. Fig. 24 visualizes an example of a Regex.

Regular Expression: `\bSS\s*[-]?\s*316\b`	
Matches:	Ignores:
✓ SS316	× pre**ss**ure of **316** bar
✓ SS 316	× more or le**ss 316**
✓ SS-316	× STRE**SS 316** MPa

Fig. 24: Example of a Regular Expression and its results.

Such Regular Expressions are useful for search tasks of targeted entities because entities follow specific conventions usually. To mention the stainless-steel alloy 316, for example, the entity used to name the alloy will often contain capital letters around the three digits “316”, like in “SS316”, “SS-316”, “316 SS”, or “AISI 316”. Regex can describe such syntactical conventions. Of course, several Regexes are necessary to cover the variety of entity synonyms and abbreviations. Meanwhile, the expression has to remain selective enough to reduce false matches. Accuracy of the Regular Expressions was ensured by a subset of articles that were traced throughout this classification step. The subset consisted of 100 randomly chosen papers, and for every alloy name, 20 manually chosen ones each. The Python code (Appendix A) found 5’137 papers with entities of “SS316”, “HT9, and “Zircaloy-4” in their texts.

5.4. Information extraction and consolidation

As described in Ch. 4.3, text mining relies on POS-tagging and entity recognition. With the lack of corpora and thus trained machine learning algorithms, information extraction from the text bodies of the papers is not possible to implement. Of the other two scientific information containing parts of articles, namely figures and tables, just tables are included in XML-files. Subsequentially, the focus shifted to information extraction out of tables. The Python code found 10’116 tables in the 5’137 files.

Topic	Entities
Dose	dose, dpa, neutron fluence, neutron flux
Temperature	temperature, °C, °F
Mechanical properties	yield strength, ultimate tensile strength, hardness
Ductility properties	uniform elongation, total elongation
Toughness	fracture toughness, J-integral
Embrittlement	upper shelf energy shift, transition temperature
Creep	creep strain, creep rate, creep compliance
Swelling	swelling rate, volume change

Fig. 25: Overview of the entities considered to describe the topic of material mechanisms in nuclear fuel-cladding applications.

Different entities relevant to material mechanisms in nuclear fuel-cladding applications were considered to be detected in order to record the content within the found tables. Fig. 25 gives an overview of the fields of material mechanisms and their entities to describe them. These entities were represented in different Regex to cover most of their synonyms, abbreviations, and unique units, while the code should provide enough selective accuracy. As in the classification before, the accuracy of this step was ensured by a subset of 100 randomly chosen papers and five manually chosen ones for every topic each. The number of appearances of the different entities in the 10’116 tables was counted and summarized in a spreadsheet. Fig. 26 outlines the structure of the spreadsheet containing the computed number of entity appearances. 285 tables contained at least one alloy and one material property entity. Strikingly, many articles in the nuclear design field were found because reactor designs are commonly summarized in tables containing information about chosen alloys, dose, temperature, and mechanical properties.

DOI	Tab_nr	Tab_text	Alloys*			Dose	Temp	YS	UTS	HV	UE	TE	F_tough	Embrit	Creep	Swelling
10.1016/j.jnucmat.2008.10.001	Table 3	Summary and literature comparison…	0	0	4	1	2	0	0	1	0	0	0	0	0	0
10.1016/j.nucengdes.2009.07.014	Table 1	Summary of design constraints for fast reactors.	0	0	0	4	14	0	0	0	0	0	0	0	0	0
10.1016/j.matchar.2017.12.038	Table 2	Hardness values reported in literature.	0	0	1	0	0	0	0	2	0	0	0	0	0	0

Fig. 26: Outline of the actual 10’116 entries long spreadsheet containing the entity appearances within the corresponding table. Temp, YS, UTS, HV, UE, TE, F_tough, Embrit stand for temperature, yield strength, ultimate tensile strength, hardness, uniform elongation, total elongation, fracture toughness, and embrittlement, respectively. Alloys* is a placeholder for the actual “SS316”, “HT9”, and “Zr4” as column names.

The ChemDataExtractor also retrieves chemical data from tables within the articles. In its information extraction step, it resorts to the trained entity recognizer to process the table layout. In this work, the information extraction had to be done manually, which increased the effort intensely. Thus, the information extraction was narrowed to the properties of SS316 only. 31 tables contained SS316, dose, and temperature entities. Especially the cited references of the found tables proved to be useful sources to build up a database of properties of SS316. Information about the purity level of the SS316 alloy or the pre-treatments, like annealed or cold-worked, were left disregarded. This reduced the dimension of the database. The material property values were gathered and occasionally converted to the units listed in Fig. 27. The resulting database contains 2’348 entries in 391 instances. It is publicly available at [50].

To summarize the amount of information gathered in the database, regression analysis was performed to fit different function types by the method of least squares. Therefore, all data instances were converted, so they became dependent on the dose received (in dpa), and the homologous temperature (in K) based on the melting point of SS316 at 1400 °C [57]. Often the information of the energy profile of the neutron flux received by the material was missing. So, the dose was calculated based on the neutron fluence values (E > 0.1 MeV) multiplied with the rough estimate of $7.8 \cdot 10^{-22}$ dpa·cm^2 [14]. These fitted functions model the behavior of the SS316 alloy depending on irradiation temperature and dose. Then, the data points and the found functions were plotted. Based on these observations and insights from literature, a diagram mapping the material’s mechanisms were compiled. Similar to the deformation-mechanism map by

Ashby, this material map might give a better overview of the mechanisms to consider at a specific temperature and dose.

Column name	Unit	Description
fluence_0.1	$n \cdot cm^{-2}$	Reported neutron fluence with E > 0.1 MeV.
fluence_1	$n \cdot cm^{-2}$	Reported neutron fluence with E > 1 MeV.
fluence_total	$n \cdot cm^{-2}$	Reported total neutron fluence.
dose	dpa	Reported dose or dpa.
irr_T	°C	Reported irradiation temperature.
test_T	°C	Reported test temperature.
YS	MPa	Reported yield strength.
UTS	MPa	Reported ultimate tensile strength.
UE	%	Reported uniform elongation or elongation at necking.
TE	%	Reported total elongation.
HV5	$kg \cdot mm^{-2}$	Reported Vickers hardness value with 5 kp test force.
HV10	$kg \cdot mm^{-2}$	Reported Vickers hardness value with 10 kp test force.
strain_rate	s^{-1}	Reported strain rate of the experiment.
t_strain_range	%	Reported total strain range of fatigue experiments.
N_fail	1	Reported load cycles until failure.
N_crack	1	Reported load cycles until crack.
J_integral	$kJ \cdot m^{-2}$	Reported J-integral.
rapture_life	h	Reported lifetime until rapture in creep experiments.
creep_strain	%	Reported creep strain of creep experiments.
density_decr	%	Reported density decrease in creep or swelling experiments.
DOI	-	Digital Object Identifier (DOI) number of the source article of the instance.

Fig. 27: Summary of the SS316 properties and their units gathered in the database.

6. Results and discussion

The data collected in the database covers the field below 60 dpa and 0.65 homologous temperature (T/T_M) decently. Fig. 28 gives an overview of the dose and temperature values of the data instances. In general, data of material properties at high temperatures and doses above 100 dpa are sparse, as also reported by the NRC [58].

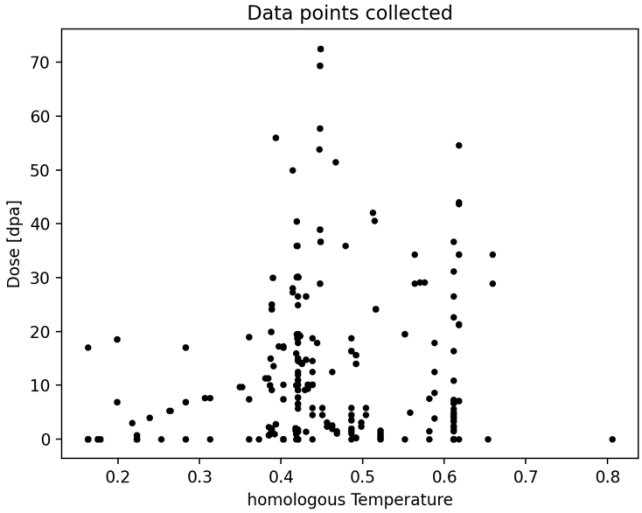


Fig. 28: Dose and temperature values of the data instances gathered.

6.1. Hardening

As explained in Ch. 3, the material mechanisms hardening, recovery, and recrystallization compete with each other when dose and temperature are elevated. Neglecting any dose, the yield strength (YS) of austenitic steels exhibits a severe drop above 0.4 T/T_M. Different approaches are trying to model that behavior. While the NRC suggests a polynomial of the fourth order [58], the National Institute of Standards and Technology (NIST) advises in its technical note the following function [59]:

$$\sigma_y(T) = a \cdot exp\left[-\frac{1}{2}\cdot\left(\frac{T}{r_3}\right)^{r_1} - \frac{1}{2}\cdot\left(\frac{T}{r_4}\right)^{r_2}\right] + d\,, \qquad \text{Eq. 14}$$

where T is the temperature (in °C), and r_1, r_2, r_3, r_4, a, and d are fitting parameters. Several studies, e.g. [24], [56], [60]–[64], investigated the behavior of the YS at certain doses and elevated temperature. As shown in Fig. 29 (b), the YS exhibits

hardening which saturates after a specific dose at low temperatures. With increasing temperature, recovery can be measured. Noteworthy, some studies report initial hardening, although the temperatures are high enough for recovery. The influence of temperature on the dose-dependent recovery is challenging to model in a single function. An exponential function can describe the hardening and saturation behavior at lower temperatures [58]. However, a third-order polynomial was chosen, although it does not describe the saturation behavior because it fitted the recovery values at higher temperatures better. So, the dose dependence of the yield strength was modeled to be:

$$\sigma_y(dpa) = k + l \cdot dpa + m \cdot dpa^2 + n \cdot dpa^3 \,, \qquad \text{Eq. 15}$$

where k, l, m, and n are fitting parameters again.

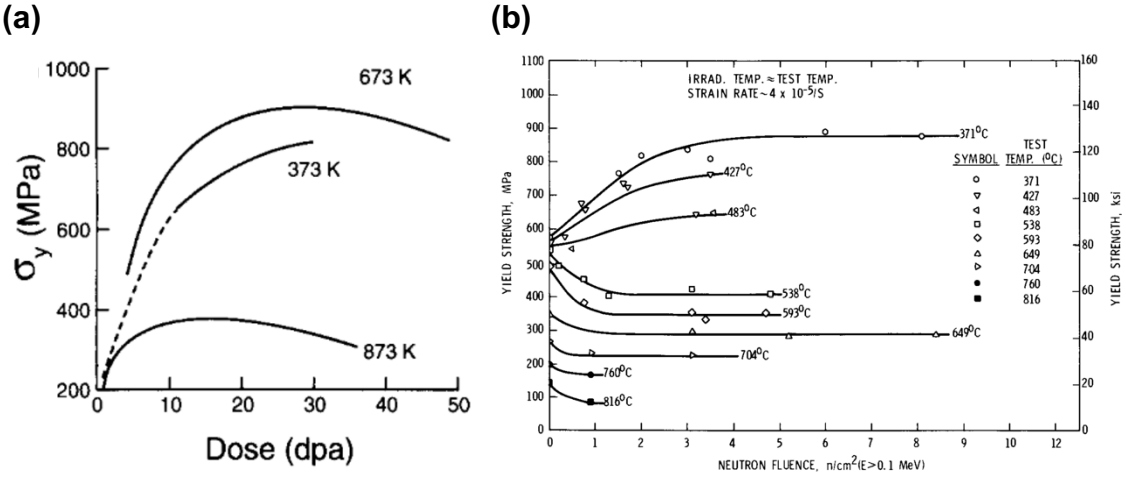


Fig. 29: Change of the yield strength by irradiation, showing initial hardening (a) and instant recovery (b) at elevated temperatures [24], [63].

Eq. 14 and Eq. 15 are then combined into the overall describing function:

$$\sigma_y(T, dpa) = p \cdot \sigma_y(T) \cdot q \cdot \sigma_y(dpa) \,, \qquad \text{Eq. 16}$$

where p and q are parameters, which need to be fitted to the dataset too. Thus, Eq. 16 is the combination of two 2D-models describing the influence of temperature and dose, respectively, on the yield strength. This approach aims to describe the experimentally measured material behavior as close as possible. All parameters in Eq. 14, Eq. 15 and Eq. 16 were computed by the Least Square method on the entire yield strength dataset. The fitted parameters of this and all the following functions are listed in Appendix E.

Fig. 30 shows the resulting yield strength function graphically. Physical reasonability of the resulting function is limited, as a second increase of yield strength above 50 dpa is unrealistic. Nevertheless, the adjusted R^2-value of the function at 0.662 is quite substantial. Although present at lower temperatures already, recovery and re-irradiation become the dominating material mechanisms after the drop of the yield strength after its maximum. The function exhibits a slight shift of the maximum YS to lower dpa values with increasing temperature.

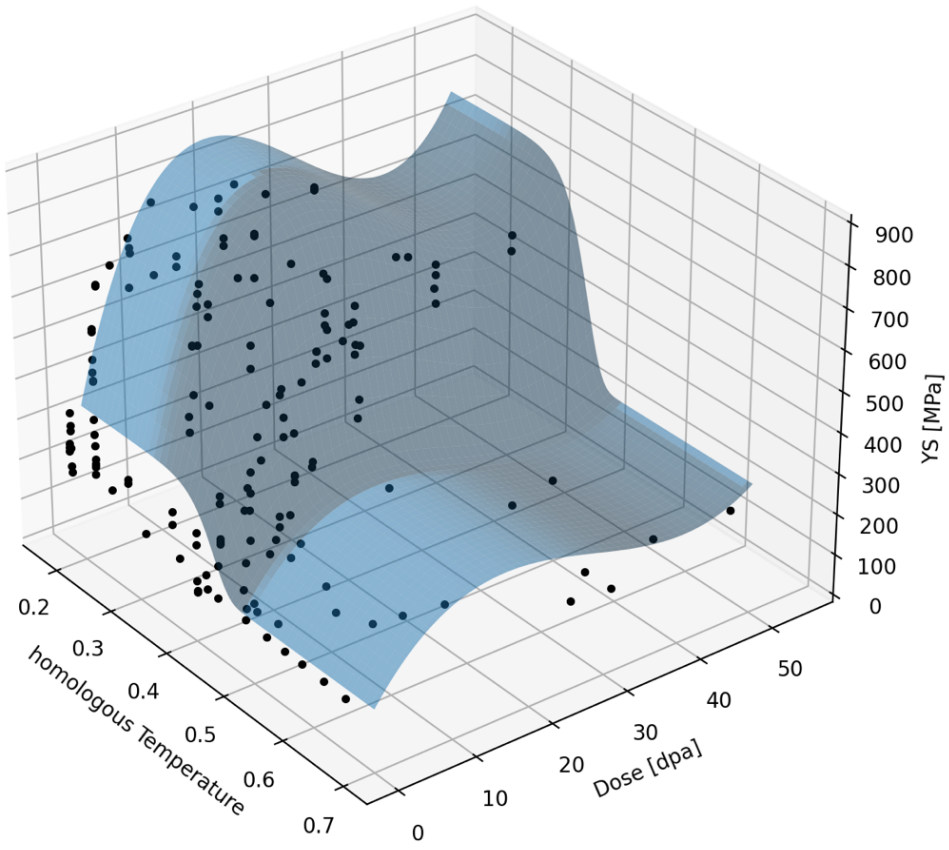


Fig. 30: Yield strength (YS) data points (in black) and the computed yield strength function σ_y(T,dpa) in blue.

Different than yield strength, a two-variate third order polynomial with temperature and dose as its variables (Appendix E) can describe the change of the ultimate tensile strength (UTS). Such a function describes the behavior reported in literature reasonably. Also, the adjusted R^2-value of 0.632 is substantial. Yet, the negative values at very high temperatures and very high doses, and the second increase of the UTS at doses above 70 dpa, are again physically unrealistic. Publicly available UTS data at very high doses and temperatures would be necessary to model a more realistic material behavior.

A comparison of the obtained YS and UTS functions, shown in Fig. 31 (b), indicates that the UTS is not as sensitive to dose than the YS. Without recovery, YS and UTS will coincide with increasing dose in certain regions. This fact is also reported by others [62] and it results in unpredictable failures at low deformations. Between 2.5 and 25 dpa and below 0.4 T/T_M, Fig. 31 (b) exhibits the region where the UTS is projected to be lower than the yield strength, which should be interpreted as the concurrence of YS and UTS.

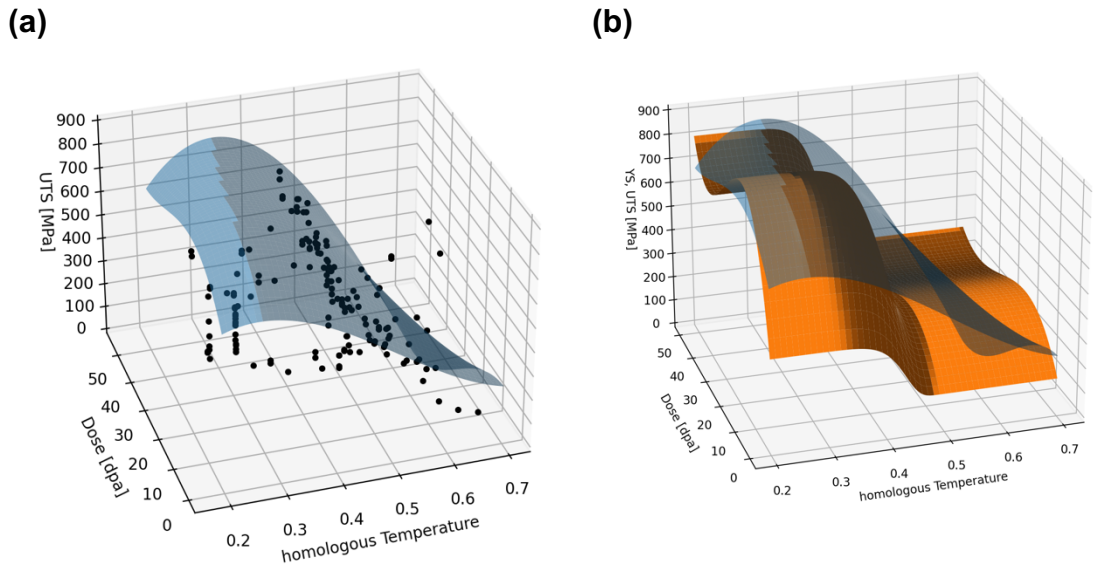


Fig. 31: (a) Ultimate tensile strength (UTS) data points and the fitted two-variant fourth-order polynomial. (b) Comparison of the two fitted functions of YS (in orange) and UTS (in blue) with each other.

6.2. Embrittlement

It was found that the best model to describe the decrease of the uniform elongation (UE) and the total elongation (TE), depending on the temperature and the dose, is a double polynomial function:

$$\varepsilon(T) = a + b \cdot T + c \cdot T^2 + d \cdot T^3 \,, \qquad \text{Eq. 17}$$

$$\varepsilon(dpa) = k + l \cdot dpa + m \cdot dpa^2 + n \cdot dpa^3 \,, \qquad \text{Eq. 18}$$

$$\varepsilon(T, dpa) = \varepsilon(T) \cdot \varepsilon(dpa) \,, \qquad \text{Eq. 19}$$

where ε stands for the elongation in both cases, and *a* till *n* are fitting parameters. This double polynomial was chosen because it results in a reasonable description of the reported material behavior. Also, its adjusted R^2-values of 0.581 and 0.461 for UE and TE, respectively, are acceptable. Different than strength, the strain

does exhibit a plateau at medium doses and temperatures [65]. As seen in Fig. 32, the function models the temperature dependence reasonably well, while it exaggerates the dose-dependent trend to another unrealistic increase. The data shows clearly that, in general, the loss of ductility, thus embrittlement, is present throughout the investigated field.

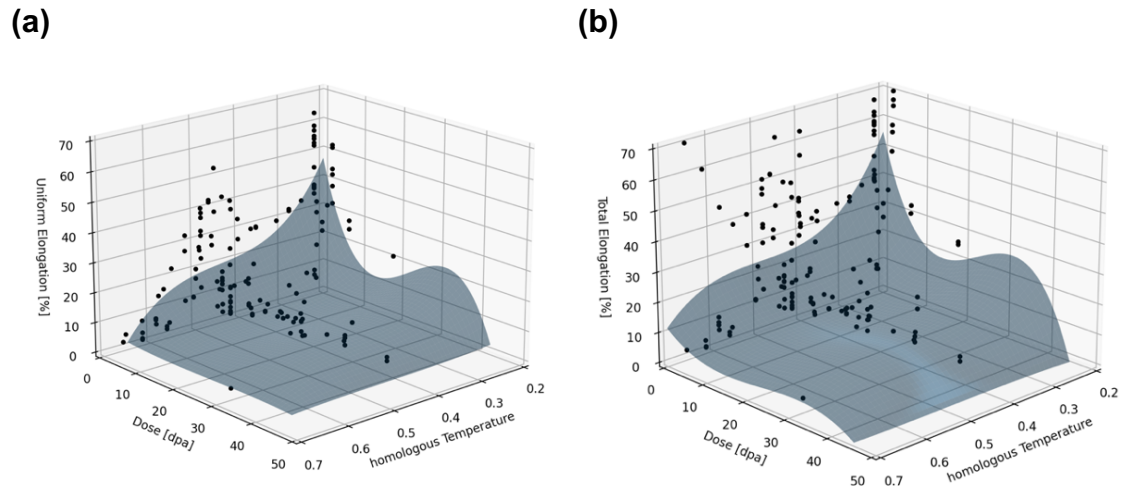


Fig. 32: Uniform elongation (UE) (a) and total elongation (TE) (b) data points and the fitted double polynomial planes (in gray).

Holmes et al. [65] summarized their work about the relationship of ductility, fluence, and temperature in a schematic diagram, as shown in Fig. 33. This figure was digitalized with the WebPlotDigitizer [66], such the boundary lines can be transferred into the material mechanism map (Fig. 38).

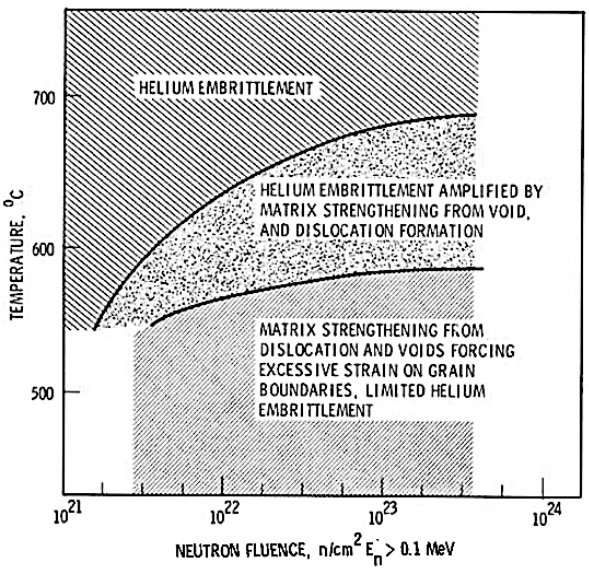


Fig. 33: Schematic diagram of various ductility degradation processes [65].

6.3. Creep and swelling

Several studies, e.g. [65], [67]–[72], investigated thermal and irradiation creep. Creep highly depends on the mechanical load or pressure acting on the material and not just on dose and temperature, as shown in Fig. 16 (b). However, irradiation creep is measurable at low doses and temperatures already. The irradiation and thermal portion on creep at high doses and high temperatures are often subject to research. Thus, the 0.4 T/T_M-line was marked in the resulting material mechanism map (Fig. 38), because both creep contributions become significant from this point onward.

Data of swelling, otherwise, often indicate an onset dose necessary, so the density decrease becomes noticeable. This behavior is also confirmed by the literature [68], [71]–[73], exemplarily shown in Fig. 34. Also, if a linear trend plane is fitted to the corresponding data in the database, an area without swelling is revealed, as seen in the contour plot Fig. 35 (b) clearly. So, the swelling boundary between 548 and 573 K in Fig. 34 was included in the material mechanism map (Fig. 38).

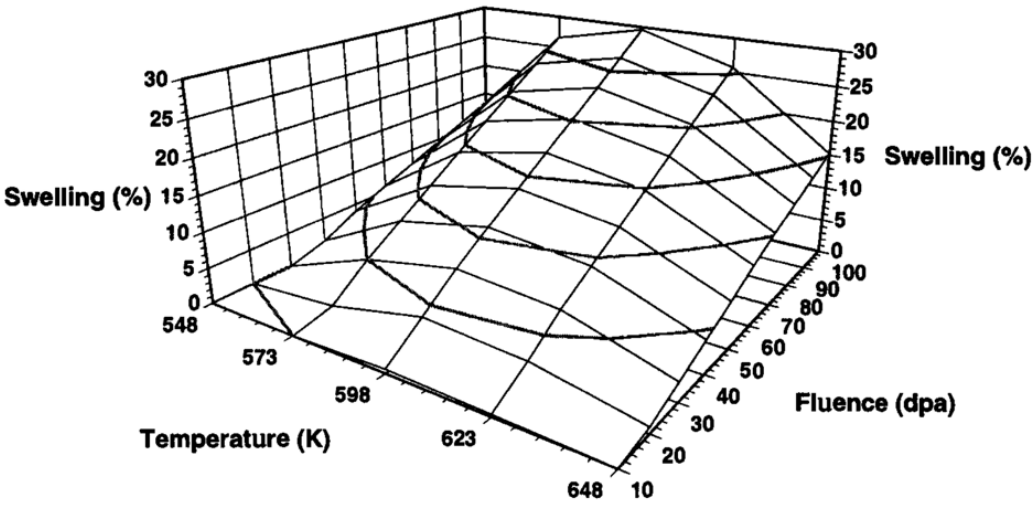


Fig. 34: Calculated dose and temperature dependence of swelling [72].

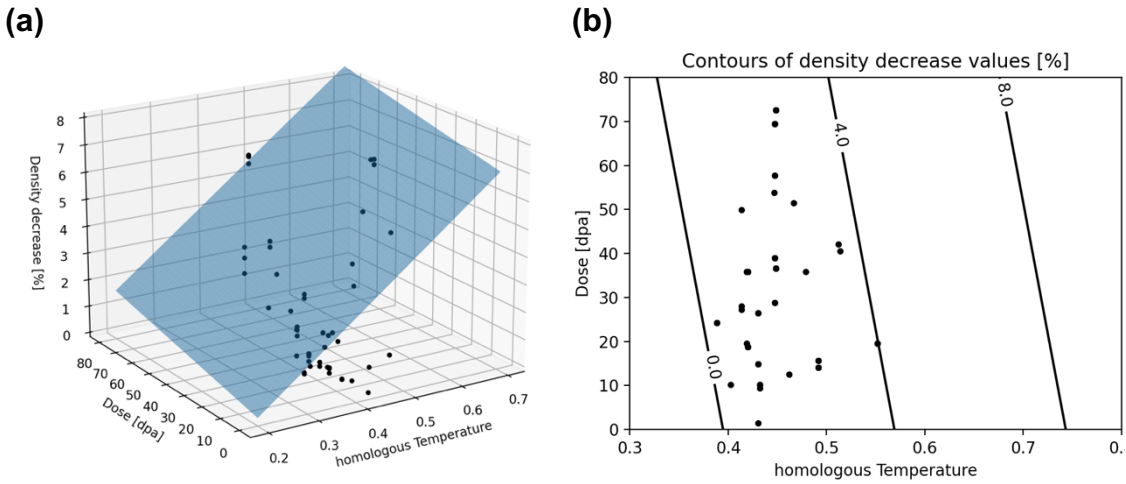


Fig. 35: Density decrease data points (in black) with the fitted linear trend plane (in blue) as a 3D-plot (a) and a contour plot (b).

6.4. Precipitation formation

Brager and Garner [71] summarized their findings of the irradiation-induced Ni_3Si (γ') formation in a diagram shown in Fig. 36. Similarly, Brager and Straalsund [74] plotted their findings of defect developments in irradiated SS316 steel in one map (Fig. 37). Fig. 37 is a perfect example of a figure with high information density. Although the aim of their study was to examen the changes of voids and bubbles, their figure also depicts information about precipitation types and conditions for precipitation formation due to causal relationships. Data points of the depicted boundaries in Fig. 36 and Fig. 37 were taken with the WebPlotDigitizer [66], so an interpolation line can be added to the material mechanism map (Fig. 38).

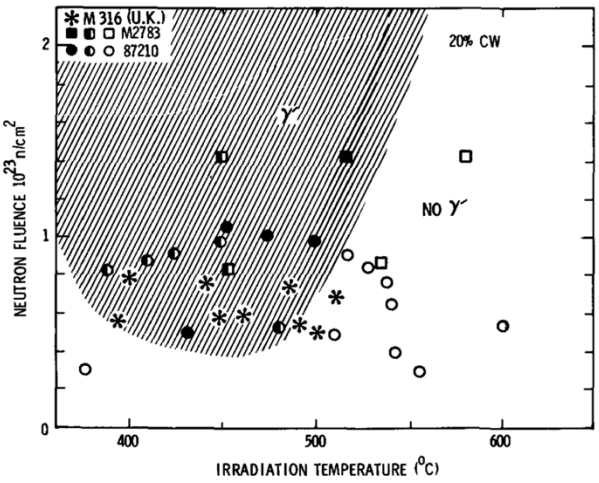


Fig. 36: Irradiation conditions for Ni_3Si (γ') formation in SS316 [71].

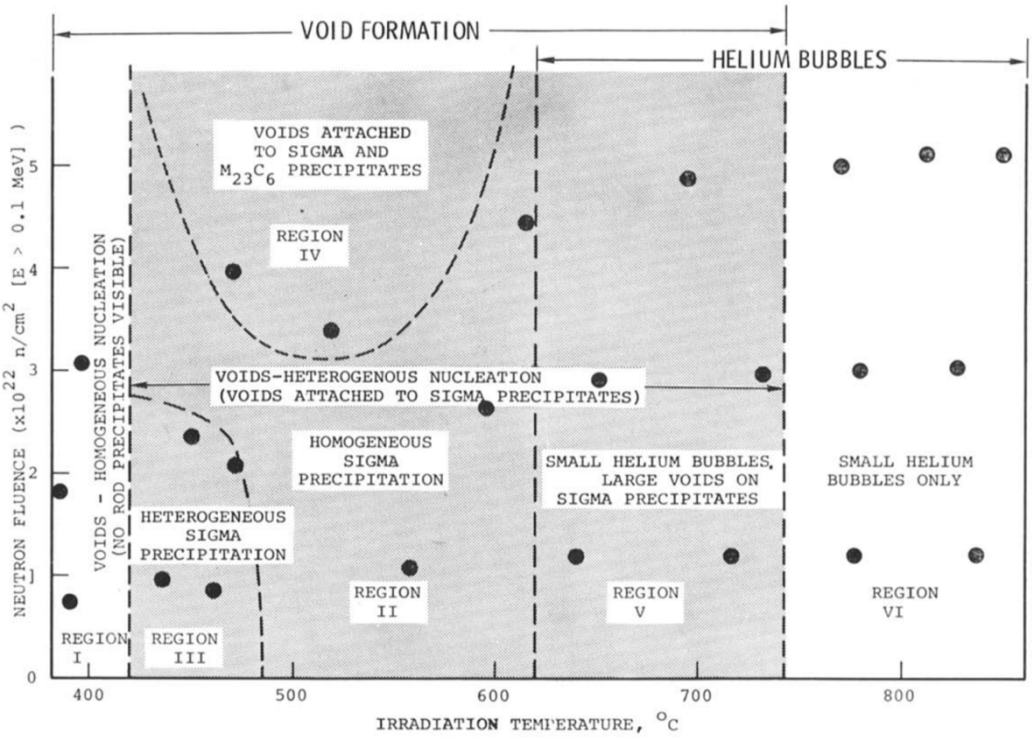


Fig. 37: Irradiation conditions for carbide and rod-shaped σ precipitation formations in SS316 [74].

6.5. Materials mechanism map

Fig. 38 is the resulting material mechanism map summarizing the information laid out in the chapters before. At the boundaries, the side of the description indicates the mechanism area where it is active. For example, swelling is present at higher temperatures and dose rates than the yellow boundary line.

For the hardening-recovery boundary, the two extreme forms of the model are represented. The vertical dashed 0.3 T/T_M-line represents instant recovery above the temperature (Model 1), while its alternative (Model 2) was constructed from the YS extreme values in [24]. Fig. 29 indicates that in both cases, the actual onset temperature for recovery is shifted by the dose anyway. The real boundary probably lies in between those two extremes. Further, the region where the YS coincides with the UTS in Fig. 31 (b) is indicated as the shaded “Brittle failure region”.

Depending on the load, irradiation creep might be present within the whole depicted area. So, a place holder at 0.4 T/T_M is added, marking roughly the moment of significant additional thermal creep. Both the trend of the data point

collection, as well as the literature, indicate a swelling onset boundary approximately in the same region. Thus, the boundary indicated in Fig. 34 was transferred into the map as well. The boundaries for carbide, σ and γ' precipitation formation, Helium embrittlement, and embrittlement by matrix strengthening were digitalized from the scientific articles [65], [71], [74].

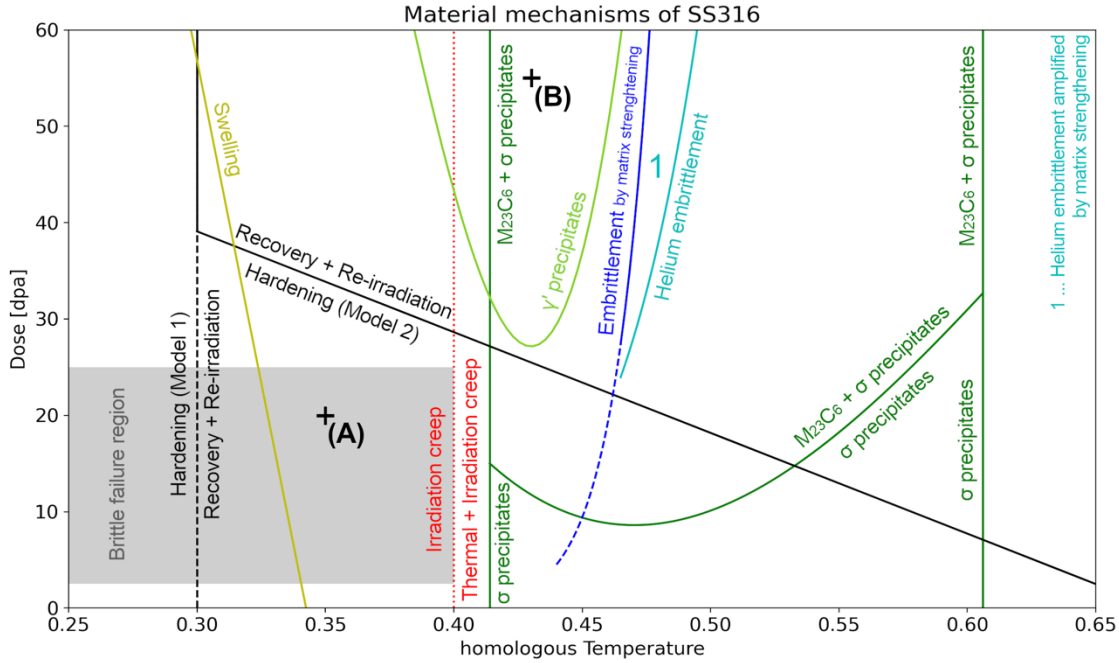


Fig. 38: Material mechanism map for SS316, giving an overview of the different mechanisms to consider at specific doses and temperatures. For example, at 0.35 T/T_M and 20 dpa (point A), one should consider swelling, hardening, the concurrence of YS and UTS, embrittlement, and irradiation creep and might neglect precipitation formations or Helium embrittlement. At 0.43 T/T_M and 55 dpa (point B), one should consider swelling, recovery and re-irradiation, embrittlement, thermal and irradiation creep, as well as the formation of carbide, σ, and γ' precipitates.

Of course, the indicated boundaries should be seen as guidelines, and exact scientific values should not be derived from them. For example, the hardening – recovery boundary comes with a big uncertainty due to contradictory models reported in the literature. Additionally, most of the indicated regions shift with parameters, which were not considered, like external load, pressure, cold-working, strain rate, or reactions with the environment. Also, the information summarized in this map would become more robust with a more significant database. Especially adding more data instances at high temperatures and doses would enhance the fitted functions modeling the empiric material behavior.

7. Conclusions

Material selection relies on the information of suitable materials, especially in nuclear applications, where material tests are complex and require special attention to safety. With their unseen service temperatures and much higher doses, the new generation of proposed nuclear reactors imposes new challenges to the known metal alloys. A more profound material selection one can make with the increasing amount and availability of the information. Peer-review scientific articles present an interesting source with high quality. Additionally, within the growing research community, it becomes increasingly harder to keep track of research progress in a specific field. Thus, the process of research and material selection could profit from automated information extraction. This thesis studies the possibilities of information extraction and information summarization of the properties of metallic alloys:

- Information extraction from a large number of scientific articles is only possible by machine learning (ML) approaches, like natural language processing.
- Those ML algorithms can process text and tables by splitting them up in smaller junks of words or word groups. These need to be characterized then, so a materials science entity, like “ultimate tensile strength” or “creep strain”, is detected as such. Therefore, the ML algorithm is trained on a collection of already labeled text, so-called corpora. For a reliable entity recognizer, several thousand articles need to be labeled manually and correctly. Such a corpus is missing in the materials science field for metal alloys properties.
- In materials science, large amounts of data are typically presented graphically in a diagram. A fully automated plot digitizer would excavate enormous amounts of material data.
- Big material databases enable the discovery of new correlations of material properties, which material scientists were not aware of yet.
- With the increasing size of the database, its clarity declines. So, the information should be summarized, for example, in a material mechanism map. This material mechanism map depicts areas of material property changes one should consider under specific environmental conditions.
- In this thesis, a material mechanism map for the austenitic steel SS316 serving as nuclear fuel-cladding material was computed as an example for information summarization.

8. Outlook on information extraction

Several databases contain information of elements, chemicals, or crystal structures, but databases of metallic alloys and steel grades are rare. Without the necessary corpora as fundaments for efficient natural language processing in the materials science field, the setup of big material databases remains laborious and unattractive. Fig. 39 summarizes the working principle and the identified challenges for fully automated information extraction. Additionally, some more challenges were identified during this thesis.

In materials science articles, the data obtained in the research is typically presented in figures and graphics. Rarely, the results, e.g. of a tensile test, are listed in a table, but in most cases in a stress-strain curve. This practice is based on the fact that figures transport much more information on a smaller visual field. Thus, automatic digitalization of data points and curves, without manual reworking, would be valuable. However, this is a more informatics related research topic. On the other hand, making the corresponding raw dataset publicly available online, simultaneously with the paper publication, would improve scientific progress too.

Noticeably, the majority of the data collected in this work was published in the 1970s. Many other papers found by the algorithm cited those articles too. The XML-files of such articles are available, but the different article parts were not marked in their files. Even worse, some XML-files just contained the article's metadata. Those are some of the reasons why the in this thesis elaborated algorithm could not find their tables automatically. Probably, these issues are consequences of scanning the paper original into a PDF and the missing sufficient converting programs.

Notably, many articles containing material properties are published on ASTM Compass and International Nuclear Information System (INIS) of the IAEA. Both platforms do not offer an API yet, like Taylor and Francis Online. Their development of such service deserves to be fostered.

The work of Swain and Cole [52] and the work of Court and Cole [75] are great examples, what could be possible based on the automated extraction of information. Court has been able to predict phase diagrams and transition temperatures of magnetic and superconducting materials. His extractor based on

the ChemDataExtractor can collect information form scientific articles for the calculation of phase diagrams of not considered compound combinations before their synthesis.

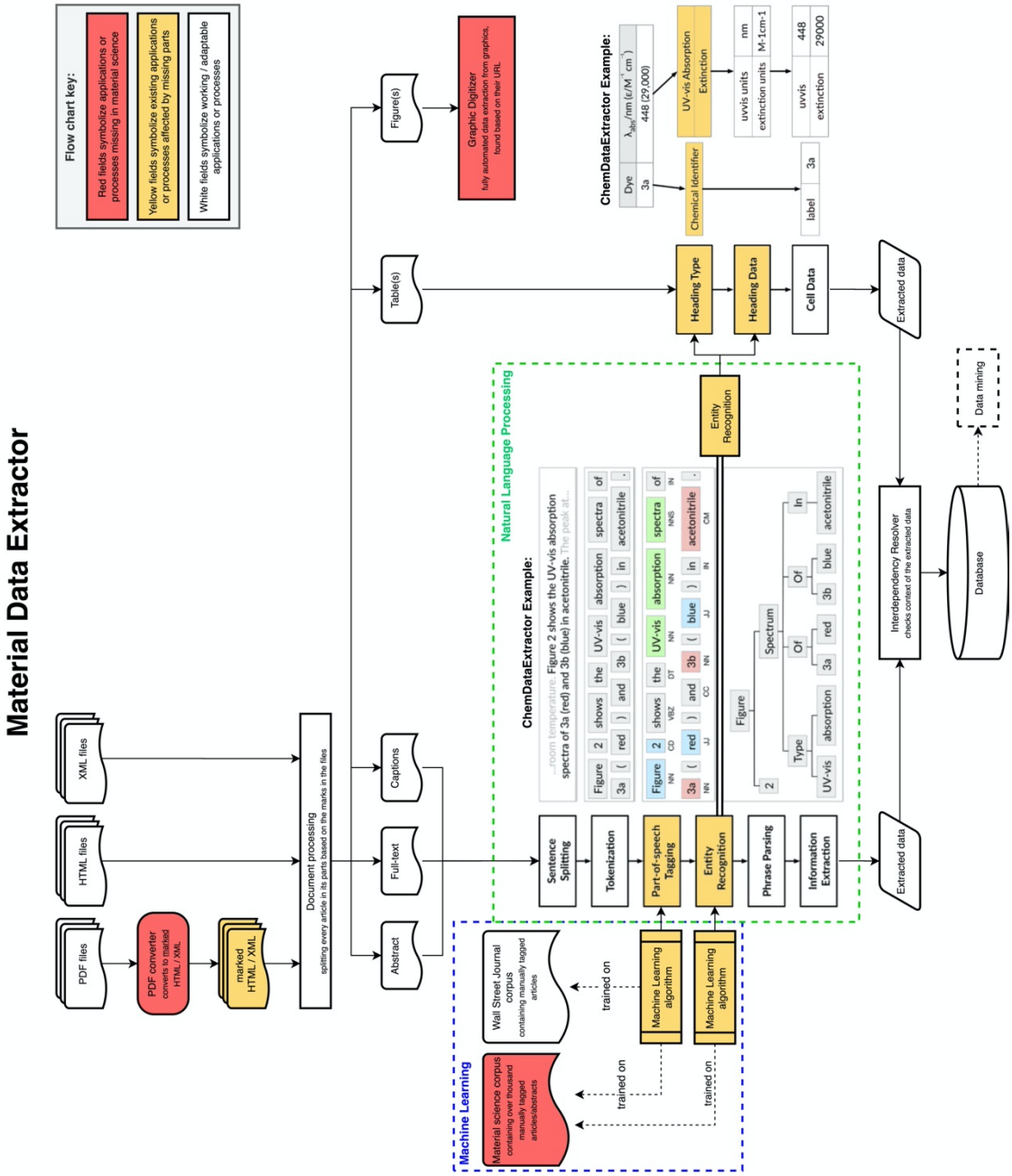


Fig. 39: Overview of the working principle of the material data extractor based on the ChemDataExtractor [52]. Red fields represent the identified parts missing in materials science to achieve fully automated information extraction.

9. References

[1] M. F. Ashby, *Materials Selection in Mechanical Design*. 3rd ed., Elsevier Butterworth-Heinemann, 2005.

[2] T. Allen, J. Busby, M. Meyer, and D. Petti, “Materials challenges for nuclear systems,” *Materials Today*, vol. 13, no. 12, pp. 14–23, 2010, doi: 10.1016/S1369-7021(10)70220-0.

[3] International Atomic Energy Agency, “IAEA Advanced Reactors Information System (ARIS).” [Online]. Available: https://aris.iaea.org. [Accessed: 02-Aug-2019].

[4] OECD Nuclear Energy Agency for the Generation IV International Forum (GIF), “Technology Roadmap Update for Generation IV Nuclear Energy Systems,” 2014. [Online]. Available: https://www.gen-4.org/gif/upload/docs/application/pdf/2014-03/gif-tru2014.pdf. [Accessed: 30-Dec-2019].

[5] P. Hosemann, D. Frazer, M. Fratoni, A. Bolind, and M. F. Ashby, “Materials selection for nuclear applications: Challenges and opportunities,” *Scripta Materialia*, vol. 143, pp. 181–187, 2018, doi: 10.1016/j.scriptamat.2017.04.027.

[6] C. R. F. Azevedo, “Selection of fuel cladding material for nuclear fission reactors,” *Engineering Failure Analysis*, vol. 18, no. 8, pp. 1943–1962, 2011, doi: 10.1016/j.engfailanal.2011.06.010.

[7] ASTM International, *A 771/A 771M – 95: Standard Specification for Seamless Austenitic and Martensitic Stainless Steel Duct Tubes for Liquid Metal-Cooled Reactor Core Components*. 2001.

[8] ASTM International, *B351/B351M-13: Standard Specification for Hot-Rolled and Cold-Finished Zirconium and Zirconium Alloy Bars, Rod, and Wire for Nuclear Application*. 2018.

[9] R. L. Klueh and A. T. Nelson, “Ferritic/martensitic steels for next-generation reactors,” *Journal of Nuclear Materials*, vol. 371, no. 1–3, pp. 37–52, 2007, doi: 10.1016/j.jnucmat.2007.05.005.

[10] P. Hosemann, S. Kabra, E. Stergar, M. J. Cappillo, and S. A. Maloy, “Micro-structural characterization of laboratory heats of the Ferric/Martensitic steels HT-9 and T91,” *Journal of Nuclear Materials*, vol. 403, no. 1–3, pp. 7–14, 2010, doi: 10.1016/j.jnucmat.2010.05.005.

[11] G. S. Brady, H. H. Clauser, and J. A. Vaccari, “Materials, Their Properties and Uses (U - Z),” in *Materials Handbook: An Encyclopedia for Managers, Technical Professionals, Purchasing and Production Managers, Technicians, and Supervisors*, 15th ed., New York: McGraw-Hill Education, 2002.

[12] C. L. Withmarsh, “Review of Zircaloy-2 and Zircaloy-4 Properties Relevant to N.S. Savannah Reactor Design,” *UC-80-Reactor Technology, TID-4500,* 17th ed., Report No. ORNL-3281, p. 70, 1962.

[13] G. Kessler, *Sustainable and Safe Nuclear Fission Energy*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[14] G. S. Was, *Fundamentals of Radiation Materials Science: Metals and Alloys, Second Edition*, 2nd ed., New York, NY: Springer New York, 2017.

[15] E. A. Kenik and J. T. Busby, “Radiation-induced degradation of stainless steel light water reactor internals,” *Materials Science and Engineering R: Reports*, vol. 73, no. 7–8, pp. 67–83, 2012, doi: 10.1016/j.mser.2012.05.002.

[16] J. L. Brimhall, H. E. Kissinger, and G. L. Kulcinski, “The Effect of Temperature on Void Formation in irradiated pure and impure metals,” in *Proceedings of radiation-induced voids in metals*, 1972, doi: 10.2172/4694493.

[17] P. J. Maziasz and C. J. McHargue, “Microstructural evolution in annealed austenitic steels during neutron irradiation,” *International Materials Reviews*, vol. 32, no. 1, pp. 190–219, 1987, doi: 10.1179/095066087790150331.

[18] P. J. Maziasz, “Overview of microstructural evolution in neutron-irradiated austenitic stainless steels,” *Journal of Nuclear Materials*, vol. 205, no. C, pp. 118–145, 1993, doi: 10.1016/0022-3115(93)90077-C.

[19] S. J. Zinkle, P. J. Maziasz, and R. E. Stoller, “Dose dependence of the microstructural evolution in neutron-irradiated austenitic stainless steel,” *Journal of Nuclear Materials*, vol. 206, no. 2–3, pp. 266–286, 1993, doi: 10.1016/0022-3115(93)90128-L.

[20] F. A. Garner and D. S. Gelles, “Neutron-Induced Swelling of Commercial Alloys at Very High Exposures,” *Effects of Radiation on Materials: 14th International Symposium (Volume II)*, ASTM STP1046, pp. 673–683, 1990.

[21] Y. Chen, “Irradiation effects of HT-9 martensitic steel,” *Nuclear Engineering and Technology*, vol. 45, no. 3, pp. 311–322, 2013, doi: 10.5516/NET.07.2013.706.

[22] P. G. Shewmon, “Radiation-induced swelling of stainless steel,” *Science*, vol. 173, no. 4001, pp. 987–991, 1971, doi: 10.1126/science.173.4001.987.

[23] P. J. Maziasz, “Temperature dependence of the dislocation microstructure of PCA austenitic stainless steel irradiated in ORR spectrally-tailored experiments,” *Journal of Nuclear Materials*, vol. 191–194, pp. 701–705, 1992, doi: 10.1016/0022-3115(92)90563-Z.

[24] G. E. Lucas, “The evolution of mechanical property change in irradiated austenitic stainless steels,” *Journal of Nuclear Materials*, vol. 206, no. 2–3, pp. 287–305, 1993, doi: 10.1016/0022-3115(93)90129-M.

[25] G. R. Irwin, “Analysis of stresses and strains near the end of a crack transversing a plate,” *Journal of Applied Mechanics-Transactions of the ASME,* vol. E24, pp. 351–369, 1957.

[26] J. R. Rice, “A path independent integral and the approximate analysis of strain concentration by notches and cracks,” *Journal of applied mechanics*, vol. 35, no. 2, pp. 379–386, 1968.

[27] A. E. Ismail, A. K. Ariffin, S. Abdullah, and M. Ghazali, “Stress intensity factors for surface cracks in round bar under single and combined loadings,” *Meccanica*, vol. 47, no. 5, pp. 1141-1156, 2012, doi: 10.1007/s11012-011-9500-7.

[28] ASTM International, *E399 – 19: Standard Test Method for Linear-Elastic Plane-Strain Fracture Toughness K_{Ic} of Metallic Materials*. 2010.

[29] International Atomic Energy Agency (IAEA), “Applicability of the leak before break concept,” Report No. IAEA-TECDOC-710, 1993. [Online]. Available: https://www-pub.iaea.org/MTCD/Publications/PDF/te_710_web.pdf. [Accessed: 03-Jan-2020].

[30] American Society of Mechanical Engineers, “ASME BPVC Section III Division 1, Subsection NB,” *ASME Boiler & Pressure Vessel Code*, 2015, doi: 10.1016/B978-1-4160-4210-5/00021-9.

[31] U.S. Nuclear Regulatory Commission, “10 CFR Part 50 - Domestic Licensing of Production and Utilization Facilities.” [Online]. Available:

https://www.nrc.gov/reading-rm/doc-collections/cfr/part050/. [Accessed: 10-Jan-2020].

[32] M. F. Ashby, “A first report on deformation-mechanism maps,” *Acta Metallurgica*, vol. 20, no. 7, pp. 887–897, 1972, doi: 10.1016/0001-6160(72)90082-X.

[33] S. J. Zinkle and G. E. Lucas, “Deformation and fracture mechanisms in irradiated FCC and BCC metals,” *Fusion Materials Semiannual Report for period ending*, DOE Office of Fusion Energy Sciences, pp. 101–125, 2003.

[34] K. Rajan, “Materials informatics,” *Materials Today*, vol. 8, no. 10, pp. 38–45, 2005, doi: 10.1016/S1369-7021(05)71123-8.

[35] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, and C. Kim, “Machine learning in materials informatics: Recent applications and prospects,” *npj Computational Materials*, vol. 3, no. 1, 2017, doi: 10.1038/s41524-017-0056-5.

[36] J. M. Rickman, T. Lookman, and S. v. Kalinin, “Materials informatics: From the atomic-level to the continuum,” *Acta Materialia*, vol. 168, pp. 473–510, 2019, doi: 10.1016/j.actamat.2019.01.051.

[37] G. J. Mulholland and S. P. Paradiso, “Perspective: Materials informatics across the product lifecycle: Selection, manufacturing, and certification,” *APL Materials*, vol. 4, no. 5, 2016, doi: 10.1063/1.4945422.

[38] National Research Council, “Materials in the New Millennium: Responding to Society’s Needs,” in *2000 National Materials Advisory Board Forum*, 2001, p. 54, doi: 10.17226/10187.

[39] Office of Science and Technology Policy (OSTP), “Materials Genome Initiative Strategic Plan,” 2014. [Online]. Available: https://www.mgi.gov/sites/default/files/documents/mgi_strategic_plan_-_dec_2014.pdf. [Accessed: 06-Jan-2020].

[40] H. L. Eiselstein and D. J. Tillack, “The Invention and Definition of Alloy 625,” in *Superalloys 718, 625 and Various Derivatives (1991)*, 1991, doi: 10.7449/1991/Superalloys_1991_1_14.

[41] A. Géron, *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*, 2nd ed., O’Reilly, 2019.

[42] T. Mueller, A. G. Kusne, and R. Ramprasad, “Machine Learning in Materials Science,” in *Reviews in Computational Chemistry*, vol. 29, A. L. Parrill and K. B. Lipkowitz, Eds. New Jersey: John Wiley & Sons, 2016, pp. 186–273.

[43] M. Jin, P. Cao, and M. P. Short, “Predicting the onset of void swelling in irradiated metals with machine learning,” *Journal of Nuclear Materials*, vol. 523, pp. 189–197, 2019, doi: 10.1016/j.jnucmat.2019.05.054.

[44] H. Zhang *et al.*, “Application of fuzzy learning in the research of binary alloys: Revisit and validation,” *Computational Materials Science*, vol. 172, 2020, doi: 10.1016/j.commatsci.2019.109350.

[45] F. Provost and T. Fawcett, *Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking*. 1st ed., Sebastopol, CA: O’Reilly, 2013, ISBN: 9781449361327.

[46] C. Shearer, “The CRISP-DM model: The New Blueprint for Data Mining,” *Journal of Data Warehousing*, vol. 5, no. 4, pp. 13–22, 2000.

[47] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed., Elsevier, 2012.

[48] A. Hotho, A. Nürnberger, and G. Paaß, “A Brief Survey of Text Mining,” *LDV Forum - GLDV Journal for Computational Linguistics and Language Technology*, 2005, doi: 10.1111/j.1365-2621.1978.tb09773.x.

[49] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python.* Sebastopol, CA: O’Reilly, 2009, ISBN: 9780596516499.

[50] R. H. Reich, “NucMatDatafication,” 2020, doi:10.5281/zenodo.4043420. [Online]. Available: https://github.com/ReneReich/NucMatDatafication. [Accessed: 22-Sep-2020].

[51] International Organization on Standardization, *ISO 32000-1: Document management - Portable document format - Part 1*. 2008.

[52] M. C. Swain and J. M. Cole, “ChemDataExtractor: A Toolkit for Automated Extraction of Chemical Information from the Scientific Literature,” *Journal of Chemical Information and Modeling*, vol. 56, no. 10, pp. 1894–1904, 2016, doi: 10.1021/acs.jcim.6b00207.

[53] C. J. Court and J. M. Cole, “Auto-generated materials database of Curie and Néel temperatures via semi-supervised relationship extraction,” *Scientific Data*, vol. 5, no. 1, 2018, doi: 10.1038/sdata.2018.111.

[54] A. Constantin, S. Pettifer, and A. Voronkov, “PDFX: fully-automated PDF-to-XML conversion of scientific literature,” in *Proceedings of the 2013 ACM symposium on Document engineering - DocEng ’13*, 2013, p. 177, doi: 10.1145/2494266.2494271.

[55] PubGenius Inc., “Typeset.” [Online]. Available: https://typeset.io. [Accessed: 24-Jul-2020].

[56] H. R. Higgy and F. H. Hammad, “Effect of fast-neutron irradiation on mechanical properties of stainless steels: AISI types 304, 316 and 347,” *Journal of Nuclear Materials*, vol. 55, no. 2, pp. 177–186, 1975, doi: 10.1016/0022-3115(75)90151-8.

[57] C. Wegst and M. Wegst, *Stahlschlüssel*, 22nd ed. Stahlschlüssel Wegst GmbH, 2010.

[58] O. K. Chopra and A. S. Rao, “Degradation of LWR core internal materials due to neutron irradiation,” *NRC publication NUREG/CR-7027*, 2010.

[59] M. Seif *et al.*, “Temperature-Dependent Material Modeling for Structural Steels: Formulation and Application,” *NIST Technical Note 1907*, Gaithersburg, MD, 2016, doi: 10.6028/NIST.TN.1907.

[60] T. Yoshitake, T. Donomae, S. Mizuta, H. Tsai, R. Strain, T. Allen, and J. Cole, "Tensile Properties of 12% Cold-Worked Type 316 Stainless Steel Irradiated in EBR-II Under Lower-Dose-Rate Conditions to High Fluence," in *Effects of Radiation on Materials: 20th International Symposium*, ed. S. Rosinski, M. Grossbeck, T. Allen, and A. Kumar, West Conshohocken, PA: ASTM International, 2001, pp. 469-486, doi: 10.1520/STP10551S

[61] R. L. Fish and J. J. Holmes, “Tensile properties of annealed type 316 stainless steel after EBR-II irradiation,” *Journal of Nuclear Materials*, vol. 46, no. 2, pp. 113–120, 1973, doi: 10.1016/0022-3115(73)90129-3.

[62] K. Farrell, T. S. Byun, and N. Hashimoto, “Deformation mode maps for tensile deformation of neutron-irradiated structural alloys,” *Journal of Nuclear Materials*, vol. 335, no. 3, pp. 471–486, 2004, doi: 10.1016/j.jnucmat.2004.08.006.

[63] R. L. Fish, N. S. Cannon, and G. L. Wire, “Tensile Property Correlations for Highly Irradiated 20 Percent Cold-Worked Type 316 Stainless Steel,” *ASTM Special Technical Publication*, no. 683, pp. 450–465, 1979, doi: 10.1520/stp38181s.

[64] C. Bailat, F. Gröschel, and M. Victoria, “Deformation modes of proton and neutron irradiated stainless steels,” *Journal of Nuclear Materials*, vol. 276, no. 1–3, pp. 283–288, 2000, doi: 10.1016/S0022-3115(99)00222-6.

[65] J. Holmes, A. Lovell, and R. Fish, "Ductility of Irradiated Type 316 Stainless Steel," in *STP529-EB Effects of Radiation on Substructure and Mechanical Properties of Metals and Alloys*, ed. J. Moteff, pp. 383-398. West Conshohocken, PA: ASTM International, 1973, doi: 10.1520/STP35463S.

[66] A. Rohatgi, “WebPlotDigitizer,” 2020. [Online]. Available: https://automeris.io/WebPlotDigitizer. [Accessed: 28-Jul-2020].

[67] D. Duncan, "Effects of Irradiation Creep on Ex-Reactor Mechanical Properties," in *Effects of Radiation on Structural Materials*, ed. J. Sprague and D. Kramer, pp. 567-577. West Conshohocken, PA: ASTM International, 1979, doi: 10.1520/STP38189S.

[68] E. Bloom and W. Wolfer, "In-Reactor Deformation and Fracture of Austenitic Stainless Steels," in *Effects of Radiation on Structural Materials*, ed. J. Sprague and D. Kramer, pp. 656-672. West Conshohocken, PA: ASTM International, 1979, doi: 10.1520/STP38195S.

[69] D. Porter and F. Garner, "Cessation of Irradiation Creep in AISI 316 Concurrent with High Levels of Swelling," in *Influence of Radiation on Material Properties: 13th International Symposium (Part II)*, ed. F. Garner, C. Henager, and N. Igata, pp. 11-21. West Conshohocken, PA: ASTM International, 1987, doi: 10.1520/STP25637S.

[70] F. A. Garner and D. L. Porter, “Irradiation creep and swelling of AISI 316 to exposures of 130 dpa at 385-400°C,” *Journal of Nuclear Materials*, vol. 155–157, Part 2, pp. 1006–1013, 1988, doi: 10.1016/0022-3115(88)90458-8.

[71] H. Brager and F. Garner, “Dependence of Void Formation on Phase Stability in Neutron-Irradiated Type 316 Stainless Steel,” in *Effects of Radiation on Structural Materials*, ed. J. Sprague and D. Kramer, pp. 207-232. West Conshohocken, PA: ASTM International, 1979, doi: 10.1520/STP38167S

[72] Y. Katoh, “Evaluation of Low-Temperature Swelling in Austenitic Stainless Steels,” in *Effects of Radiation on Materials: 18th International Symposium*, ed. R. Nanstad, M. Hamilton, F. Garner, and A. Kumar, pp.

783-793. West Conshohocken, PA: ASTM International, 1999, doi: 10.1520/STP13903S.

[73] J. I. Bramman *et al.*, “Void Swelling and microstructural changes in fuel pin cladding and unstressed specimens irradiated in DFR,” *NRC publication NUREG/CR-7128*, 1977.

[74] H. R. Brager and J. L. Straalsund, “Defect development in neutron irradiated type 316 stainless steel,” *Journal of Nuclear Materials*, vol. 46, no. 2, pp. 134–158, 1973, doi: 10.1016/0022-3115(73)90131-1.

[75] C. J. Court and J. M. Cole, “Magnetic and superconducting phase diagrams and transition temperatures predicted using text mining and machine learning,” *npj Computational Materials*, vol. 6, no. 1, p. 18, 2020, doi: 10.1038/s41524-020-0287-8.

Appendix A: Python code collecting XML files

```python
import requests
import numpy as np
from bs4 import BeautifulSoup
from elsapy.elssearch import ElsSearch
from elsapy.elsclient import ElsClient


def get_DOI_from_Springer(query: str, api_key: str) -> list:
    springer_url = "http://api.springernature.com/metadata/pam"
    page_step = 50
    doi_list = []
    url = springer_url + f"?q={query}&s=1&p=1&api_key={api_key}"
    soup = BeautifulSoup(requests.get(url).text, "xml")
    total = soup.find("total").get_text()
    print("Total articles found: " + total)
    for page in list(np.arange(1, int(total), page_step)):
        url = springer_url + f"?q={query}&s={page}&p={page_step}&api_key={api_key}"
        soup = BeautifulSoup(requests.get(url).text, "xml")
        doi_list.extend([doi.get_text() for doi in soup.find_all("dc:identifier")])
    return doi_list


def get_DOI_from_Elsevier(query: str, api_key: str) -> list:
    config = {"apikey": api_key, "insttoken": ""}
    client = ElsClient(config['apikey'])
    client.inst_token = config['insttoken']
    doc_srch = ElsSearch(query, 'sciencedirect')
    doc_srch.execute(client, get_all=True)
    doi_list = []
    for element in doc_srch.results:
        doi_list.append(element.get('dc:identifier'))
    print("Total articles found: " + str(len(doi_list)))
    return doi_list


def download_XML(publisher: str, doi_list: list, dst_path: str, api_key: str):

    if not publisher == "Elsevier" or publisher == "Springer":
        return print("Please choose the publisher correctly: 'Elsevier' or 'Springer'")
    count = 0

    for doi in doi_list:
        file_path = doi.replace('/', '%')
        if publisher == "Elsevier":
            headers = {
                'Accept': 'application/xml',
                'X-ELS-APIKey': api_key
            }
            url = 'https://api.elsevier.com/content/article/doi/' + str(doi)
            r = requests.get(url, stream=True, headers=headers)
        else:
            springer_url = 'https://spdi.public.springernature.app/xmldata'
            r = requests.get(springer_url + f'/jats?q={doi}&api_key={api_key}')
        with open(dst_path + f'/{file_path}.xml', 'wb') as path:
            path.write(r.content)
        count += 1
        print(f"Downloaded: {count} of {len(doi_list)} XML files")
```

Appendix B: Python code classifying files

```
import os, re
import xml.etree.ElementTree as ET
import pandas as pd


def classify_file_by(directory_path: str, entities: dict):
    data_frame = []
    count = 0
    file_list = os.listdir(directory_path)
    for file in file_list:
        print(f"File now being classified: {file}")
        try:
            tree = ET.parse(os.path.join(src, file))
        except:
            continue
        root = tree.getroot()
        xml_string = str(ET.tostring(root, encoding='utf8', method='xml'))
        for expression in entities.values():
            if re.search(expression, xml_string) and file not in data_frame:
                data_frame.append(file)
                count += 1
                print(f"Entities found in {count} papers")
    df = pd.DataFrame(data_frame)
    df.to_excel("Files_classifications.xlsx")
    return print('Classifications exported to "Files_classifications.xlsx"')


src = "/directory full of XMLs/path/"
material_list = {
    '316SS':    'SS\s*[-]?\s*316|316\w*\s*[-]?\s*SS|316\w*\s*[-]?\s*(stainless)?\s*[-]?\s*steel|'
                '(stainless)?\s*[-]?\s*steel\s*[-]?\s*316|AISI\s*[-]?\s*\w*\s*[-]?\s*316',
    'HT9':      '(?:HT|Ht|ht)\s*[-]?\s*9',
    'Zr4':      '([Zz]ircal{1,2}oy|[Zz]r)\s*[-]?\s*4',
                }
classify_file_by(src, material_list)
```

Appendix C: Python code finding entities in tables

```python
import re, os
from bs4 import BeautifulSoup
import pandas as pd
import numpy as np

def find_entities_in_tables(dirctory_path: str, file_list, entities: dict):
    data_frame = []
    columns= ['DOI', 'Tab_nr', 'Tab_text'] + [key for key in entities.keys()]

    count = 0
    for file_name in file_list:
        with open(os.path.join(dirctory_path, file_name)) as file:
            soup = BeautifulSoup(file)
        tables = []
        tables.extend(soup.find_all("table-wrap"))
        tables.extend(soup.find_all("ce:table"))
        doi = file_name.replace('%', '/').replace('.xml', '')
        print(f'File now being searched: {doi}')
        for table in tables:
            instance = [doi]
            try:
                instance.append(table.find("ce:label").getText())
            except:
                try:
                    instance.append(table.find("label").getText())
                except:
                    instance.append(np.NaN)
            try:
                instance.append(table.find("ce:simple-para").getText())
            except:
                try:
                    instance.append(table.find("p").getText())
                except:
                    instance.append(np.NaN)
            table_string = str(table.text)
            for expression in expressions.values():
                instance.append(len(re.findall(expression, table_string)))
            data_frame.append(instance)
        count += 1
        print(f'file {count} of {len(file_list)} processed')
    df = pd.DataFrame(data_frame, columns=columns)
    df.to_excel("Entities_in_tables_v3_1.xlsx")


file_list = pd.read_excel("Papers_containing_alloys.xlsx")
src = "/Users/Rene/Dropbox/Master Thesis/Paper Population"
expressions = {
    '316SS':    'SS\s*[-]?\s*316|316\w*\s*[-]?\s*SS|316\w*\s*[-]?\s*(stainless)?\s*[-]?\s*steel|(stainless)?\s*[-]?\s*steel\s*[-]?\s*316|AISI\s*[-]?\s*\w*\s*[-]?\s*316',
    'HT9':      '(?:HT|Ht|ht)\s*[-]?\s*9',
    'Zr4':      '([Zz]ircal{1,2}oy|[Zz]r)\s*[-]?\s*4',
    'Dose':     '\b[Dd]ose\b|\bdpa\b|\bDPA\b|n\/cm\^*(<sup loc="post">)*2|[Nn]*(eutron)*\s*[-]?\s*([Ff]luence|[Ff]lux)|[Ff]luence|[Ff]lux',
    'Temp':     '[Tt]emperature|[Tt]emp\.*|°C|°F|\bT\b|\bK\b',
    'YS':       '[Yy]ield[-]?\s*[-]?\s*[Ss](?:trength|tress)|\bYS\b',
    'UTS':      '(?:[Uu]ltimate\s*[-]?\s*(?:[Tt]ensile)*|[Tt]ensile)\s*[-]?\s*[Ss](?:trength|tress)|\bUTS\b|\bTS\b',
    'HV':       '[Hh]ardness|[Hh]ardening|\bHV\b|\bHRC?\b',
    'UE':       '[Uu]niform\s*[-]?\s*(?:[Pp]lastic)?\s*[-]?\s*(?:[Ee]longation|[Ss]train)|\bUE\b',
```

```
    'TE':       '(?:[Tt]otal|[Uu]ltimate)\s*[-]?\s*(?:[Pp]lastic)?\s*[-
]?\s*(?:[Ee]longation|[Ss]train)|(?:[Ee]longation|[Ss]train)\s*[-]?\s*at\s*[-
]?\s*(?:[Bb]reak|[Ff]racture)|\bTE\b',
    'F_tough':  '[Ff]racture\s*[-]?\s*[Tt]oughness|[Jj]\s*[-
]?\s*[Ii]ntegral|k?J\/m\^*(<sup loc="post">)*2|MPa\s?.?\s?m\^*(<sup loc="post"
>)*1\/2|[Ff]racture',
    'Embrit':   '(?:[Dd]uctile\s*[-]?\s*[Bb]rittle)?\s*[-
]?\s*[Tt]ransition\s*[-
]?\s*(?:[Tt]emperature|[Tt]emp\.*|\bΔT\b|\bTT\b)|(?:[Uu]pper\s*[-
]?\s*[Ss]helf\s*[-]?\s*[Ee]nergy|USE)\s*[-
]?\s*[Ss]hift|\bΔ?USE\b|[Ee]mbrittlement|[Ii]mpact\s*[-]?\s*[Tt]est|Charpy',
    'Creep':    '[Cc]reep\s*[-
]?\s*(?:[Rr]ate|[Ss]train|[Cc]ompliance)|[Tt]ensile\s*[-
]?\s*[Ss]train|[Ss]train\s*[-]?\s*[Rr]ate|[Cc]reep',
    'Swelling': '[Ss]welling\s*[-]?\s*[Rr]ate|[Ss]welling'
}
find_entities_in_tables(src, file_list.values, expressions)
```

Appendix D: Python code plotting the material mechanism map

```python
import matplotlib.pyplot as plt
import numpy as np


def linear_function(x, coeff, displ):
    return coeff * x + displ


def power_function(x, exp, coeff, displ):
    return coeff * x ** exp + displ


def gamma_precipitation(temperature):
    params = [129811, -145706, 53297, -6270.4]
    polynomial: float = 0
    order = len(params) - 1
    for i, p in enumerate(params):
        polynomial += p * temperature ** (order - i)
    return polynomial


def carbide_precipitation(temperature):
    params = [-3721.5, 7061.9, -4173, 796.32]
    polynomial: float = 0
    order = len(params) - 1
    for i, p in enumerate(params):
        polynomial += p * temperature ** (order - i)
    return polynomial


def hardening(temperature):
    params = [-104.57, 70.463]
    return linear_function(temperature, *params)


def swelling(temperature):
    params = [-1338.5, 458.4]
    return linear_function(temperature, *params)


def he_embrittlement(temperature):
    params = [14.8, 2E6, 0]
    return power_function(temperature, *params)


def matrix_strengthening(temperature):
    params = [32.665, 2E12, 0]
    return power_function(temperature, *params)


plt.rcParams.update({'font.size': 14})
max_dpa = 60
fig, ax = plt.subplots()
ax.plot(np.linspace(.3, .8, 200), gamma_precipitation(np.linspace(.3, .8, 200)), "#85d22e")
ax.plot(np.linspace(.414, .606, 200), carbide_precipitation(np.linspace(.414, .606, 200)), "g-")
ax.vlines(.414, 0, max_dpa, colors="g")  # lower sigma-precipitation limit
ax.vlines(.606, 0, max_dpa, colors="g")  # upper sigma-precipitation limit
```

```
58. ax.plot(np.linspace(.3, .8, 200), hardening(np.linspace(.3, .8, 200)), "k-
    ")  # hardening (model 2) limit
59. ax.vlines(.3, hardening(.3), max_dpa, colors="k")
60. ax.vlines(.3, 0, max_dpa, colors="k", linestyles="dashed")  # hardening (model
     1) limit
61. ax.fill_between(np.linspace(.15, .4, 200), 2.5, 25, facecolor='#CFCFCF')  # br
    ittle failure region
62. ax.plot(np.linspace(0, .8, 200), swelling(np.linspace(0, .8, 200)), "y-")
63. ax.vlines(.4, 0, max_dpa, linestyles="dotted", colors="r")  # thermal creep li
    mit
64. ax.plot(np.linspace(.465, .8, 200), matrix_strengthening(np.linspace(.465, .8,
     200)), "b-")
65. ax.plot(np.linspace(.44, .465, 200), matrix_strengthening(np.linspace(.44, .46
    5, 200)), "b--")
66. ax.plot(np.linspace(.465, .8, 200), he_embrittlement(np.linspace(.465, .8, 200
    )), "c-")
67. ax.set_title("Material mechanisms of SS316")
68. ax.set_xlabel('homologous Temperature')
69. ax.set_ylabel('Dose [dpa]')
70. ax.set_xlim(.25, .65)
71. ax.set_ylim(0, max_dpa)
72. plt.show()
```

Appendix E: Parameters of the fitted functions

Eq. 14	a	r_1	r_2	r_3	r_4	d
YS(T)	0.318876984	18.5265562	11.7588423	0.418179520	3.59603477	0.187349721

Eq. 15	k	l	m	n
YS(dpa)	2.80000188e+03	5.61932706e+02	-2.35774901e+01	2.74088924e-01

Eq. 16	p	q
YS(T,dpa)	5.10570844e-01	5.0573754e-01

Eq. 17	a	b	c	d
UE(T)	8.49337847e-02	-4.85010749e-01	1.06781565e+00	-8.01985255e-01

Eq. 18	k	l	m	n
UE(dpa)	1.92637231e+03	-2.41587712e+02	1.22918216e+01	-1.78411064e-01

Eq. 17	a	b	c	d
TE(T)	6.28821166e-01	-3.29584472e+00	6.70478283e+00	-4.56338573e+00

Eq. 18	k	l	m	n
TE(dpa)	2.73066196e+02	-2.89639199e+01	1.42399236e+00	-2.04339625e-02

$$UTS(T, dpa) = a + b(T) + c(dpa) + dT^2 + e(T)(dpa) + f(dpa)^2 + g(T)^3 + h(T)^2(dpa) + i(T)(dpa)^2 + j(dpa)^3$$

a	b	c	d	e
1.23190167e+02	4.64797535e+03	1.30728828e+01	-1.18952194e+04	6.36538533e+01
f	g	h	i	j
-8.61480333e-01	7.81075997e+03	-1.20157338e+02	5.18521072e-01	7.08209060e-03