The Use of Quantitative Databases in Aladin, an Alloy Design System

Ingemar Hulthage, Michael D. Rychener,
Mark S. Fox, and Martha L. Farinacci*

CMU-RI-TR-85-19 3

Intelligent Systems Laboratory
The Robotics Institute
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

*Alcoa Laboratories
Alcoa Center, Pennsylvania 15069

September 1985

This research was supported by Alcoa.

Table of Contents

1 Introduction 2
2 Overview of Aladin 2
3 The Aladin Database 5
  3.1 Phase Diagram Representation 5
  3.2 Microstructure Representation 7
4 Coupling of Symbolic and Numeric Computation 8
  4.1 Properties and Microstructure 9
  4.2 Regression Among Known Alloys 10
  4.3 Multidimensional Constraints 11
5 Conclusion 12
6 References 13

Abstract

Aladin is an expert system that assists metallurgists in the design and evaluation of new alloys. This paper gives a preliminary overview of the architecture of Aladin with emphasis on the coupling of symbolic and numeric computations. We briefly describe the knowledge representation of Alloys, Phase diagrams and Microstructure information. We explain how qualitative reasoning is used to define the variables to be dealt with quantitatively and how regression among known alloys is used. We describe multidimensional constraints and how they are used to deal with interaction between variables, to implement a least commitment strategy and how they at the same time provide criteria for search termination and backtracking.

1 Introduction

The Aluminum Alloy Design Inventor (Aladin) is an expert system that aids in the design of new alloys. The system uses traditional, rule based, qualitative methods as well as quantitative calculations. Our approach is to model the design process as a collaboration among several different bodies of expertise: composition in terms of elements to be added; properties of known alloys; internal structure, ranging from solid phases and microstructure down to the crystal lattice structure; and thermo-mechanical processing of the metal. The inference mechanism is a variant of hypothesize-and-test. The present implementation combines the Carnegie Representation Language (CRL) [2, 3] and OPS5 [4, 5] along with Lisp. A database of metallurgical knowledge and data on known alloys is built in the form of CRL schemata. The execution of the system is controlled by OPS productions and most of the data produced in a specific run is represented by OPS working memory elements. The code for interfacing CRL and OPS and for most of the calculations described in this paper is written in Lisp and the system runs in Common Lisp [6] the VAX/VMS operating system. The system is, at the time of writing, under intensive development and we can therefore only give a preliminary and incomplete description of Aladin. We will therefore, in this paper, focus on some issues relating to the coupling of symbolic and numerical computations. Specifically, we will discuss the use of a database of numerical data and the use of quantitative relationships in generating and evaluating alloy hypotheses both qualitatively and quantitatively. A more detailed description of the Aladin system will be given in a forthcoming article [1]. This report was submitted to The Workshop on Coupling Symbolic and Numerical Computing in Expert Systems held at Boeing Computer Services, Bellevue Washington, August 27-29, 1985.

2 Overview of Aladin

The problem of designing an alloy is defined by a list of criteria that should be satisfied by the new alloy. In general the definition of an alloy, as opposed to its functional criteria, is composed of two parts:

1. **Composition:** The elements constituting the alloy.

2. **Process**: The processes by which the alloy is produced.

This information is represented in a schema in the database.

```
{{ X
    INSTANCE: alloy
    MAJOR-ALLOYING-ELEMENT: Mg
    MAJOR-ALLOYING-PERCENT: (2.0 3.0)

    ULTIMATE-TENSILE-STRENGTH:
    TENSILE-YIELD-STRENGTH:
    COMPRESSIVE-YIELD-STRENGTH:

    PROCESS-METHODS: preheat-spec-x extrude-spec-x
            age-spec-x
    .
    .    }}
```

The target alloy is defined by a schema in a similar way, and the design criteria are represented as constraints on the slot values, which in CRL are attached as meta information to the slots. It is important to note that with few exceptions the constraints on the properties, as well as their values, are given in numerical form.

A careful analysis of the alloy design task suggests that the reasoning involves four interrelated spaces:

1. **Property Space**: The multidimensional space of all property variables.
2. **Composition Space**: The multidimensional space of elementary composition percentages.
3. **Processing Space**: The space of alternative manufacturing processes, qualitatively and quantitatively described.
4. **Structure Space**: The space of alternative micro-structures qualitatively and quantitatively described.

Much of the expert knowledge can be described as qualitative and quantitative heuristic rules that map changes in one space to changes in another. An other important source of knowledge is the database of known commercial and experimental alloys with a quantitative description of their microstructure, composition, manufacturing and properties. In addition there is also a rnota space with knowledge about the design process itself, and each of the four spaces has a qualitative and a quantitative sub level.

The problem of discovering a new alloy may be viewed as a search in the variable space defining an alloy. The initial state is a "root" alloy whose functional behavior is to be modified by altering the composition and processing. Each alteration is to be rated in reference to the functional criteria defined in the target alloy schema. Again, in specifying alterations and in evaluating them, quantitative methods must be employed.

Theoretically the design should begin at the most abstract level which is the structure space. Once the structure is decided upon, then a composition and processing would be selected to give the required structure and meet the property targets. In reality, the design of an alloy usually takes place in all spaces simultaneously or alternatingly in an opportunistic way, depending upon which space contains the most useful or most reliable knowledge at each stage. The design process is highly constrained since decisions made in one space will in general constrain the search in the others.

Aladin breaks down the problem of designing an alloy into subproblems and subtargets. For each subproblem a set of hypotheses is generated that describes possible ways to meet the subtarget. When one hypothesis has been picked a new subproblem is considered and hypotheses relating to it are generated. This cycle proceeds in the following manner. When a hypothesis is selected a hypothesis evaluation phase is entered. During this phase the target properties are estimated. In the next phase the estimated properties are compared to the constraints which define the design targets and if one or more differences are detected hypotheses are generated to eliminate one of these differences. In general a number of hypotheses to meet a certain property now exist, but none have been selected for further consideration. A hypothesis selection phase is therefore entered. Each hypothesis is associated with a credibility that roughly estimates the probability that the hypothesis will be a part of the final design and the selection among the descendant hypotheses is made on basis with the highest credibility is selected marking the completion of one iteration in the cycle.

The cycle described here continues in each space and at all levels until all the initial constraints have been satisfied or no more can be satisfied. The default sequence is to activate on a qualitative level first the micro structure knowledge source then composition and processing after which everything is repeated on the quantitative level. However, the power of the system is substantially increased by switching between spaces and levels opportunistically and by backtracking among the successive hypotheses.

3 The Aladin Database

In the design of new alloys, researchers depend heavily on their understanding of existing alloy systems, standard production methods, and the observed effects of composition, treatments and structural variations on key properties. The amount of knowledge required to successfully develop new materials is so great that individuals often can not complete the task alone. The knowledge bank of alloy and metallurgical information is therefore an important part of the Aladin system. The alloy database has information on known alloys including representations of product types, applications, alloy properties and production methods. The database of metallurgical knowledge has representations of phases and phase diagrams, phase transitions, microstructure and crystal structure.

3.1 Phase Diagram Representation

Phase diagrams represent much metallurgical knowledge and are central for most reasoning about metallurgy. Phase diagrams are therefore crucial for the design of alloys and Aladin has a representation for phases and phase diagrams. A phase diagram is a set of phases that describes a system and a phase is a region of a phase diagram. Therefore, an important fact associated with every phase is the domain in the space of variables (e.g., temperature and percentages of additions) in which the phase is stable. The construction of a representation for the domains of phases is explained in this section.

A good choice of phase diagram representation facilitates the retrieval of information from phase diagrams. The representation described here has been chosen to fulfill the following criteria

- A simple algorithm should exist for determining if an arbitrary point is located in the domain of the phase.
- A simple algorithm should exist for finding the neighboring phases and the points of contact with those.
- A simple algorithm should exist for finding the intersection of a phase boundary and an arbitrary value of some variable, e.g., at what temperature does the phase transition between alpha-phase and alpha + beta-phase take place for the composition 99.5% Al and 0.5% Si?
- It must be possible to represent any phase diagram with desired accuracy.
- The representation must be general enough to handle phase diagrams of any dimension, i.e., any number of variables.

The representation chosen can be described as follows. Consider a D-dimensional space. Any D +1 points in the space defines a simplex, i.e., in two dimensions any three points defines a triangle and in three dimensions any four points defines a tetrahedron, etc. Any domain with linear boundaries, i.e., lines in two dimensions and planes in three dimensions, can be exactly represented by a set of simplexes. If the boundaries are curved the domain can be approximately represented in the same way and the accuracy can be made arbitrarily good by taking a large enough number of simplexes.

A set of D + 1 -multiples of integers is associated with each particular phase of a phase system. Each multiples refers to a simplex as described above and the integers addresses reference points in the space of phase variables associated with the phase system as a whole.

The boundary, if any, between two phases can be found by searching for points that are used to define simplexes in both phases. When the boundary between two phases is known, the intersection of this boundary with, for example, a line of constant composition (in two dimensions) can also be found.

In terms of CRL schemata the representation of a the Aluminum-Lithium system is typical:

```
{{ Al-Li-system
      SUB-REGIONS: liquid-Li liquid-beta-Li beta-Li liquid-Al4-IJ9 Al4-Li9 theta-Li
            liquid-theta-Li liquid-alpha-Li alpha-theta-Li alpha-Li
      INSTANCE: phase-system
      COMPONENTS: Al Li
      EUTECTIC-TEMPERATURE: 601
      EUTECTIC-COMPOSITION: 28
      REFERENCE-POINTS: ((0 0) (G60.1 0) (601 14.6) (601 28) (0/15.8) (601 45.8) (718 50) (0 55)
                  (522 55) (0 69) (180 69) (522 69) (522 69.5) (180 98.5) (0 100)
                  (180 100))}}
```

For each of the phases (listed in slot sub-regions above) there is a schema with specific information including a definition of the region of the phase in terms of the reference points in the Al-Li-system schema:

```
{{ alpha-Li
      REGION-OF: Al-Li-system
      INSTANCE: solid-solution
      CRYSTAL-STRUCTURE: all-faces-centered-cubic
      PEARSON-SYMBOL: CF4
      SPACE-GROUP: Fm3m
      REGION: (1 2 3)}}
```

3.2 Microstructure Representation

The term microstructure will here be used for alloy features between the levels of phases and macroscopic structures exclusively, while the word structure includes both microstructure and phases. An ideal phase is produced under perfect equilibriums but in practice the absence of equilibrium introduces structural features. Hence, microstructure can be defined to be the configuration in three dimensional space of all types of non-equilibrium defects [7]. The object of the microstructure representation in Aladin is to classify and quantify the microstructure of alloys in order to facilitate the formulation of rules that relates the microstructure to the properties of alloys.

The basic elements of microstructure are the lattice defects and interfaces. The Aladin database defines four types of microstructural elements: vacancy, dislocation, grain boundary and dispersoid.

The characterization of a microstructure includes information about the the types of microstructural elements present as well as their size, shape, orientation and distribution. The representation of alloys includes slots that defines the microstructure

---

```
{{ X
     INSTANCE: alloy
     MAJOR-ALLOYING-ELEMENT: Li
     MAJOR-ALLOYING-PERCENT: (2.0 3.0)
     .
     .
     MICROSTRUCTURE: two-phase-dispersion-X
     MICROSTRUCTURAL-ELEMENTS: precipitate-X
     .
     .    }}
```

---

and the schemata two-phase-dispersion-X and precipitate-X contains, for example, volume fraction and size information. This information is important in order to determine the impact of microstructure on properties.

4 Coupling of Symbolic and Numeric Computation

In the course of the design of an alloy, quantitative information and quantitative relationships are crucial and related to discrete variables such as the choice of thermo-mechanical processes and their sequencing. This makes Aladin different from the majority of expert systems that deal with a finite set of operators and an enumerable set of states. In the Aladin system, the design starts on a qualitative level where little numerical calculation is done. On this level qualitative decisions on the microstructure and processing are made and the alloying elements are chosen. Most of these decisions have to be quantified, e.g., the percentages of alloying elements have to be determined. A least commitment strategy is used to specify the values of these variables. Instead of assigning single values to the variables defined on the qualitative level, bounds on the values consistent with the target properties are determined. The bounds take the form of linear constraining equations that are tailored to the specific case at hand and to available data. Once all the constraints are obtained, it is determined if they form a system of constraint that have a feasible region. If not backtracking to the qualitative level is neccesarry. If a feasible region exists the design could stop or some strategy could be used to pick one or more points in the allowed region. A utility function, reflecting for example cost, could be constructed and a point minimizing this function sought. Alternatively a point in the center of the region can be picked to minimize the risk of a mistake or several sample points can be chosen, according to statistical principles, for laboratory confirmation of the properties.

4.1 Properties and Microstructure

All alloy properties (except weight) depend on the microstructure. However, there are many properties for which the structure dependence is weak or relatively easy to calculate. This is true about the class of properties that to a high degree of accuracy only depends on the volume fraction of the phases present in the microstructure. Properties such as density, modulus, electrical and thermal conductivity and coefficient of expansion belong to this class. Within a single phase, linear interpolation over the composition is usually sufficient to calculate most properties and the interpolation coefficients are in many cases available from the literature. For a multiple phase microstructure, a property can be approximated with an average of the phases present weighted by the volume fraction of them. This is likely to be a good approximation if the structure dependence, is weak. However, the fact that knowledge about the volume fraction is neccesarry is a complication. To calculate the volume fraction of phases in a system with N alloying elements, access to the corresponding N-dimensional phase diagram is needed and few phase diagrams beyond $N = 3$ are known. In addition, if meta stable phases are present, details of the physical processes that produced the microstructure are in principle needed. In the absence of any of this information additional approximations have to be introduced, such as using two dimensional phase diagrams instead of multidimensional and empirical formulas describing the dynamics of meta stable phases. A linear equation that constrains the design variables involved can be generated in this manner. It is clear that a metallurgist makes many intelligent choices in order to construct a quantitative model using techniques similar to the ones described here and an expert system must be able to reproduce this reasoning.

For properties with strong structure dependence, like fracture toughness, several decisions about the types of microstructural elements present, the global structure characteristics and fracture mechanisms must be made. Since some microstructure concepts are not numerical, this reasoning is primarily symbolic. The structure decisions later restrict the choices of valid composition and process decisions.

4.2 Regression Among Known Alloys

Aladin utilizes data on known alloys in order to model the dependence between properties and design variables. Trends and estimates of properties can be discerned intuitively by studying plots of data and, more formally, by regression analysis. The problem a human expert encounters here is the volume of data and the multitude of analyses that could be performed in each case. Although regression analysis is well known and easy to program, some method to use it in conjunction with heuristic knowledge needs to be devised.

On a qualitative level, Aladin generates hypotheses by searching the database for alloys that have properties close to the target properties and by identifying design variables that, with some probability, can be assumed to bring a certain property of an alloy towards the target. Conversely, the effect on all the properties of a hypothesis generated in this way, or by some other heuristic method, can be estimated by looking at alloys of similar design. The spread in the value of a property among alloys so selected establishes a range in which the new alloy is likely to fall.

On the quantitative, level numerical data is considered explicitly and regression is used to deduce the sought quantity. However, before any computation can commence, the data to be used must be carefully selected. Only alloys of the same temper, with similar microstructure and with the same phases should be used. It is also an advantage if the composition doesn't differ much; in general the closer the data points are to the alloy under consideration the more reliable the analysis will be. We also think it is important to distinguish between interpolation and extrapolation. Whenever possible we choose points that define a hull with the new alloy in the interior. In these regressions we consider all the composition variables at all times but only one property at a time and the regression produces a constraining equation that will become a part of a system of equations.

In order to generate a quantitative hypothesis about how to reach a certain property target, the corresponding qualitative hypothesis, previously generated with or without resort to the alloy database, is considered. Typically this hypothesis states that a certain element should be added or a processing method must be used. The amount of the element to add, or the setting of process parameters is now determined through inverse regression. The effect of this hypothesis on other properties can now be evaluated through regression and the design process can proceed with the next unsatisfied target.

4.3 Multidimensional Constraints

The target properties generate constraints on the design variables through quantitative relationships between properties and the variables. Since a variable will, in general, be constrained by more than one property target, the bounds on one variable will depend on the bounds on others. The constraints, therefore, have to be described as a region in a multidimensional space of design variables. This region is defined by a system of linear equations. The objectives of the representation of constraint used in Aladin are:

- to describe a domain in a space of quantifiable design variables in which certain design targets are satisfied,
- to allow a determination of the feasibility of the constraints,
- to give a basis for a final commitment to a specific structure, composition and processing.

The result of the qualitative plan is used to determine the variables to be constrained. The percentages of the elements selected in composition space are obvious variables, but others, like process temperatures and sizes of microstructural elements, are also important. Interactions between properties can also be useful.

The formulas for properties with simple structure dependence yields constraining equations, and constraining equations for other properties can be obtained by regression in the alloy databases, as described above. Some variables, like temper, are not easily quantifiable, but have an indirect impact on the generated constraints. The temper information, for example, is used to select the alloys in the regression. Another source of constraining equations are the phase diagrams. Several heuristic rules involve phase boundaries and solubility limits and these restrictions can be expressed as constraining equations.

Once the variables and constraining equations are determined, they are most conveniently represented as matrices. There is a extensive literature on systems of linear equations. The simplex method, for example, can be used to determine if the constraint region is empty or to find points in the region.

5 Conclusion

The coupling of symbolic and numerical calculations in Aladin has been described. Symbolic reasoning is used to develop an abstract plan for alloy design. This abstract plan contains decisions about the general microstructural features of the alloy to be designed as well as the alloying elements present and the processing methods used. These qualitative decisions serve as constraints on the quantitative decisions that are made later. The coupling of qualitative and quantitative decisions may be direct. For example a decision to add an element X is an abstraction of the decision to add a certain amount of element X. The relationships may be indirect, as is the case when a decision to create certain microstructural particles constrains composition decisions and influences the interpretation of phase diagrams. Quantitative decisions are made using a least commitment approach, and well established methods, such as regression analysis and linear programming, are used to model this approach.

6 References

1. Fox M.S., Hulthage I., Rychener M. and Farinacci M. *Overview of Aladin, Aluminum Alloy Design Inventor* Technical Report, Robotics Institute, Carnegie-Mellon University, To be published.

2. Wright J.M., Fox M.S. and Adam D. *SRL/1.5 Users Manual.* Technical Report, Robotics Institute, Carnegie-Mellon University, 1984

3. *Knowledge Craft, Beta release Notes.* Carnegie Group inc. 650 Commerce Court, Station Square, Pittsburgh, PA 15219

4. Forgy C.L. *OPS5 User's Manual.* Department of Computer Science, Carnegie-Mellon University, 1981.

5. Allen B.P., *ISLops: An extended OPS5 for use with SRL1.* Technical Report, Robotics Institute, Carnegie-Mellon University, 1984.

6. Steele G.L. Jr. *Common LISP: The Language.* Digital Press 1984.

7. Hornbogen E. *On the Microstructure of Alloys.* Acta Metall. vol 32, No. 5, p. 615, 1984.