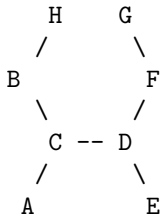
Problem 3 Molecular Constraint Propagation (20 points)

We can (approximately) model a molecule as being a branching chain of atoms. For example, in the following picture, the atoms are labelled by letters and a link indicates a chemical bond between atoms. Each of the bonds, e.g. A-C, B-C, C-D, would have a characteristic length that is known a priori. Furthermore, the angles between adjacent bonds, e.g. A-C-B, B-C-D, A-C-D, C-D-F, would also be known. Note that since both the angles and lengths are fixed and known, we also know the distances between atoms at the ends of adjacent bonds, e.g. (A,B), (A,D), (B,D).



Note that (in three dimensions) fixing these distances and angles still allows rotation about the bonds. So, for example, the group H-B-C-A can rotate around the C-D bond, as can the group G-F-D-E. So, the distances between B and F or B and E are *not* fixed. Similarly, the group F-G can rotate about the D-F bond and the group B-H can rotate around the B-C bond.

Atoms also have fixed, known, radii. So you can think of the molecule as an assembly of balls connected by sticks. That is, part of one atom may not occupy the same space as part of another, which means that the distances between the centers of a pair of atoms must be greater than the sum of their radii.

One important method for discovering the three-dimensional structure of molecules is through NMR (nuclear magnetic resonance). One type of NMR measurement can give us distances between SOME (but not all) pairs of atoms in a molecule. So, for example, let's assume that we discover from such an experiment that the distance between atoms A and G is 4 angstroms and that the distance between G and H is 3.5 angstroms.

Given a molecule whose atom-bond network is known and so the distances and angles between bonded atoms are known, we would like to use the set of distance ranges obtained through NMR and determine the positions of all the atoms.

To make the problem concrete, we imagine working on a three-dimensional grid with a fixed resolution, say 0.1 angstroms. The positions of all atoms will be limited to being at grid points. We assume we know the grid positions of three of the molecule's atoms in the grid, e.g. B, C, D in the example above; this eliminates arbitrary translations and rotations of the molecule. Then, our goal is to find grid positions for all the remaining atoms so that all the distance constraints are satisfied (taking into account a little slop introduced by the limited grid resolution). This problem when formulated in this way, is a classic CSP problem, with variables, values and pairwise constraints between variables.

Part A (5 points)

Assuming we use the atoms as variables and the grid points as values, indicate the different types of constraints required in this problem. Use only as many boxes as necessary.

1.

2.

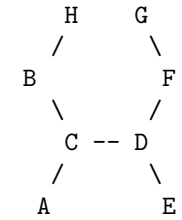
3.

4.

5.

Part B (5 points)

For each of the following pairs of atoms in the sample molecule given earlier and repeated below, indicate all the types of constraints that hold between them. Use the numbers you assigned to the constraint types in Part A. Don't list redundant constraints. If there are no constraints, write "None".



Atom Pair	Constraint	Constraint	Constraint	Constraint	Constraint
A,B					
B,C					
B,F					
B,G					
G,H					

Part C (10 points)

We will now consider a very simplified situation in which there are only four atoms: A, B, C and D and each of them have only two legal grid positions, which we will write as: A1, A2 (for atom A), B1, B2 (for atom B), C1, C2 (for atom C) and D1, D2 (for atom D).

You are told that the **only** variable-value combinations that satisfy the geometric constraints are as follows:

- **A-B**: A1-B1, A2-B2
- **A-C**: A1-C1, A2-C2
- **B-D**: B2-D2
- **C-D**: C2-D2
- **B-C**: No constraint.
- **A-D**: No constraint.

No other combination of variable values satisfies the geometric constraints.

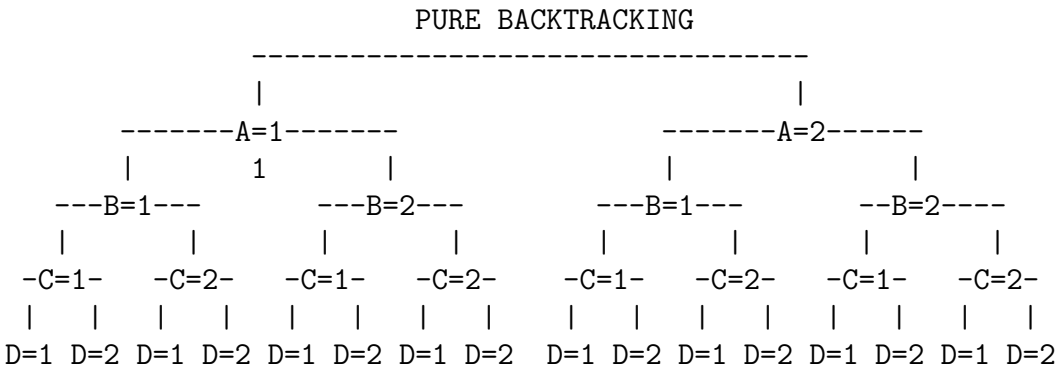
Let's say that "an assignment is generated" every time a variable in the problem gets a new (tentative) assignment.

You are to assume that tentative assignments are made before consistency is checked or forward checking is initiated.

Further assume that the variables are examined in alphabetical order and the values in numerical order.

On the next page, we ask you to solve this problem using pure backtracking and also by using backtracking with forward checking. Show your work by labeling the nodes in the search trees (each node corresponds to an assignment) with the order in which the nodes are generated in each of the two cases. *Write the number right below each assignment node, as shown for the first assignment, that of A=1.*

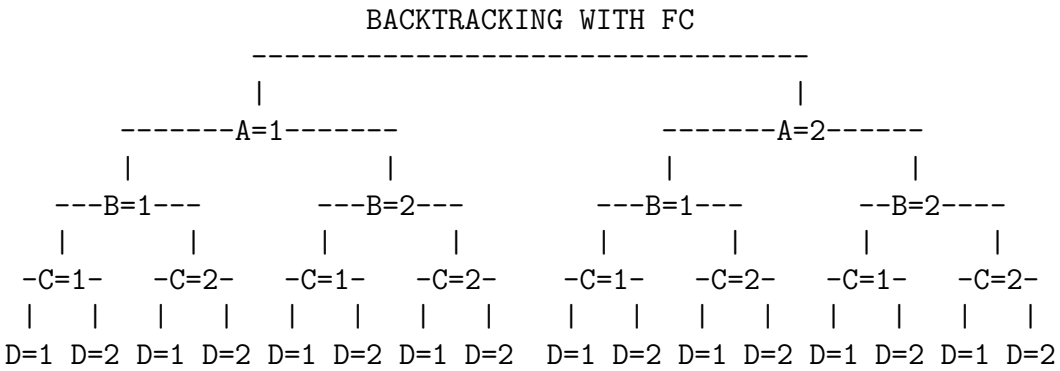
C.1 (5 points)



Using pure backtracking, how many assignments are generated during the search for the first correct, complete assignment (that is, what is the value of the final label you placed in the search tree).

Answer:

C.2 (5 points)



Using backtracking with forward checking, how many assignments are generated during the search for the first correct, complete assignment (that is, what is the value of the final label you placed in the search tree).

Answer: