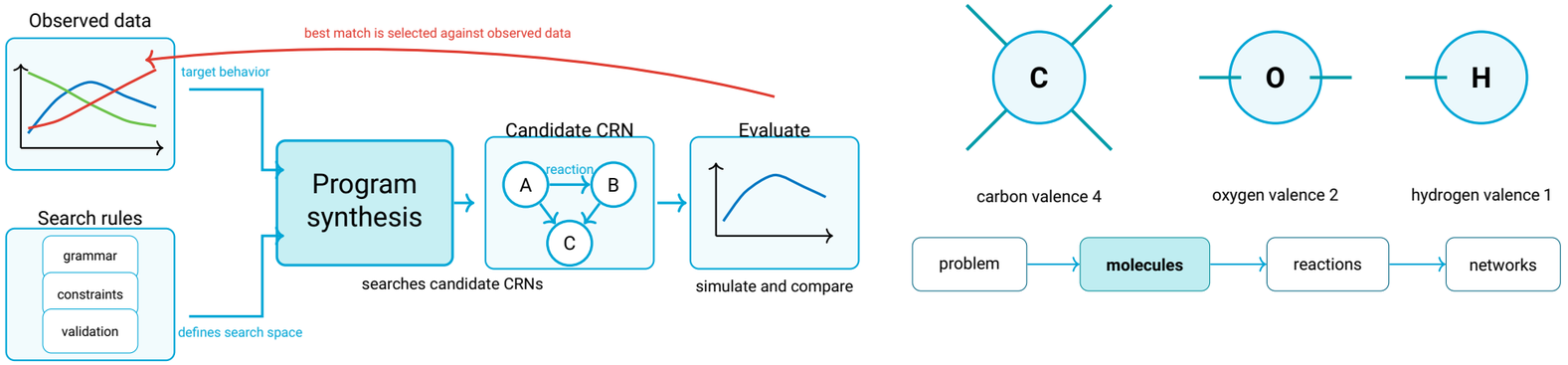
From Post-Validation to Grammar-Level Pruning: Valence Constraints for Molecule Synthesis

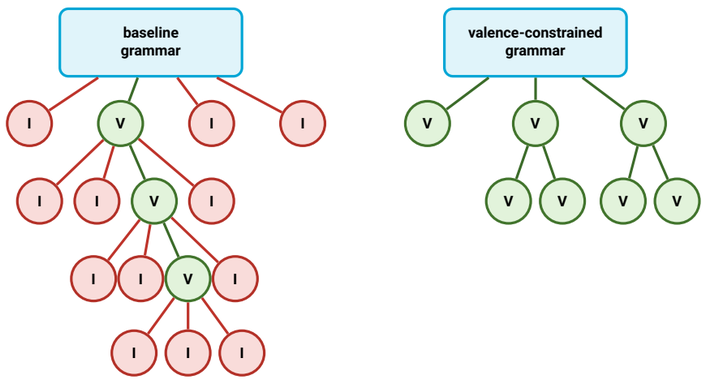
Midas van Veen | CSE3000 Research Project, Delft University of Technology

1 Background



2 Problem Statement

- Baseline BFS expands many partial molecules that are already chemically impossible.
- Those invalid branches still consume queue space and runtime before post-validation rejects them.
- Valence constraints should prune these branches at grammar expansion time.



Research Question

How can atom valence be encoded as a grammar constraint within a program synthesis framework for CRNs?

- Is the set of valid molecules preserved?
- What is the runtime trade-off?

3 Method

Baseline rule expansion

$$\text{chain} \longrightarrow \text{atom branches bond chain}$$

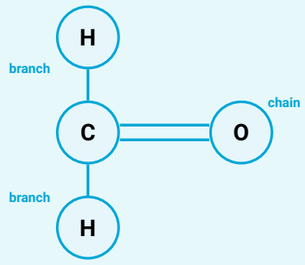
Valence-constrained rule expansion

$$\text{chain\_i} \longrightarrow \text{atom\_v branches\_\{v-i-o\} bond\_o chain\_o}$$

- i: incoming bond order
- v: current atom valence
- o: outgoing bond order

Example

$$\text{chain\_0} \longrightarrow [\text{C}]\ \text{branches\_2 bond\_2 chain\_2}$$



4 Results

Runtime: Water

Depth	Count	Old (ms)	New (ms)	Speedup
1	0	0.12	0.25	0.48×
2	0	0.14	0.27	0.53×
3	0	0.16	0.28	0.58×
4	2	0.30	0.36	0.83×
5	3	0.39	0.38	1.03×
6	4	0.54	0.44	1.22×
7	9	1.16	0.65	1.78×
8	16	2.35	1.02	2.30×
9	25	4.18	1.59	2.63×
10	36	6.93	2.41	2.87×

Runtime: Methane

Depth	Count	Old (ms)	New (ms)	Speedup
1	0	0.16	0.81	0.20×
5	4	0.50	1.21	0.41×
6	7	0.69	1.60	0.43×
8	258	17.11	44.27	0.39×
9	5063	457.19	1290.02	0.35×

Candidates to Target

Problem	Old count	New count	Count ratio
Methane ($O_2 + H_2O$)	3	5	1.67×
Estherification (full molecule goal)	61	370	6.07×

5 Discussion and Conclusion

- Ringless outputs match the legacy accepted molecule set.
- Water reaches a 2.87× speedup by depth 10.
- Methane remains slower because grammar overhead dominates.
- Grammar-level pruning helps when invalid expansions dominate the search.
- Ring closure still needs a non-local constraint.

Delft University of Technology