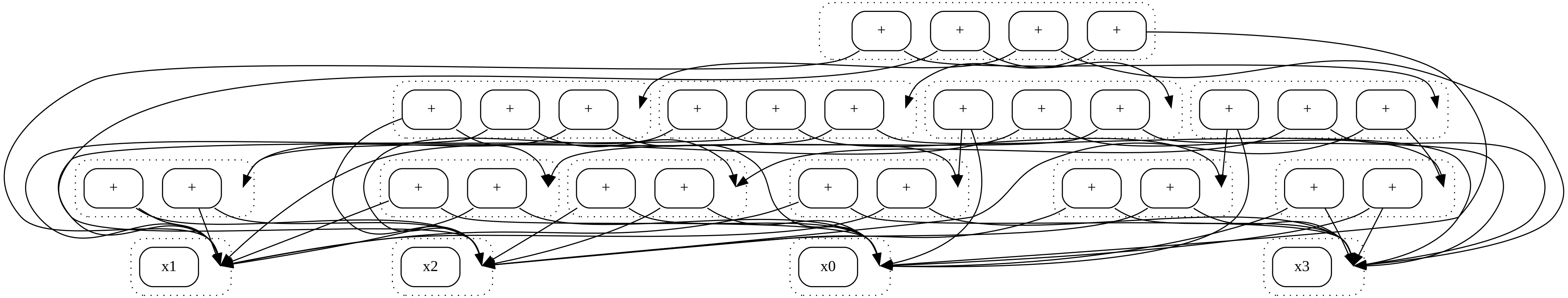
# Omelets need Onions

## E-graphs Modulo Theories via Bottom Up E-Matching

Philip Zucker (Draper Labs)

# Motivation: AC Sucks

- The Eqsat Paradox
- $(x_1 + (x_2 + \ldots (x_{N-1} + x_N) \ldots ))$
- #e-classes: $2^N - 1$



# E-Graphs Modulo Theories

- Can we bake in domain specific smarts?
    - Not Just AC: polynomial, linear, sets
- Spirit Guide: EMT ~ SMT - SAT
- E-graph sharing makes confusing 😵‍💫

# Tease Apart the Roles

E-graphs are:

- Term banks `add_term : t -> term -> unit`
- Term finders `match : t -> pat -> subst list`
- Equality stores `assert_eq : t -> term -> term -> unit`

# Term Banks Modulo Theories

- Rigid baked in "nice" theories.
- Interning by structural normalization
    - Smart constructors
        - Ex: $x + 0 \to x$

```
def add(x,y):
  return x if y == 0 else hashcons(("+", x, y))
```

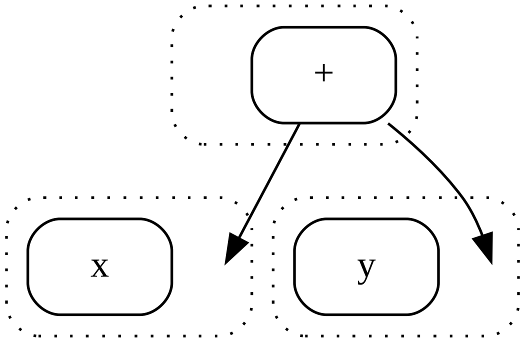
# Term Banks Modulo Theories

| | |
|---|---|
| `add_term : t -> term -> unit` | ✅ |
| `match : t -> pat -> subst list` | ❓ |
| `assert_eq : t -> term -> term -> unit` | ❓ |

# Pattern Matching

- Implicit terms
  - Consider pattern `?x + 0`

```
add_term((x + 0) + y)
```



# Top Down E-matching

- Scan termbank for term roots
- #substitutions depends on theory
    - Factor $F$ at each theory node of pattern

| Theory | Pattern | Theory Factor $F$ |
|---|---|---|
| ADT | $cons(X, Y) =^? cons(1, nil)$ | 1 |
| E-Graph | $foo(X, Y) \in^? \{foo(e_1, e_2), bar(e_2)\}$ | $\|eclass\|$ |
| MultiSet 1 | $[X, Y, Z] =^? [1, 2, 3]$ | (#Vars)! |
| MultiSet 2 | $X + Y =^? [1, 2, 3]$ | #Partitions |
| Linear | $X + Y =^? 42$ | $\infty$ |

# KEY IDEA: Bottom Up E-matching

- E-match *over the term bank*, not on term
  - `match : term -> pat -> subst list`
  - `match : termbank -> pat -> subst list`
- Bind variables by traversing term bank
  - Ex: $foo(bar(X), Y) \to biz(X)$
- Optimizations

```
for X in terms:
  for Y in terms:
    lhs = foo[bar[X], Y]
    if lhs in terms:
      rhs = biz(X)
      add_equality(lhs, rhs)
```

# Bottom Up E-matching Plays Nicer with Theories

| | TD | BU |
|---|---|---|
| Cost | $O(TF^d)$ | $O(T^V d \ln(T))$ |
| $foo(foo(foo(foo(X))))$ | 😢 | 😃 |
| $foo(X, Y, Z, W, V, U)$ | 😃 | 😢 |

- Pareto frontier for simplicity-power
  - Grounds fast
  - Only needs canonizer, not expander / unapply

# Tying the Knot

| | |
|---|---|
| `add_term : t -> term -> unit` | ✅ |
| `match : t -> pat -> subst list` | ✅ |
| `assert_eq : t -> term -> term -> unit` | ❓ |

# Q: What does the Union Find do?

```
type t
type id
val is_eq : t -> id -> id -> bool
val fresh : t -> id
val canon : t -> id -> id
val assert_eq : t -> id -> id -> unit
```

- But not only a union find presents this interface!

# KEY IDEA: Structured E-ids

- Alternative names: Semantic e-ids, *Values*
- *E-graphs are Models* (for a partial logic)
  - $\downarrow t$ and $t_1 = t_2$
- Replace union find with theory specific *extensible* canonizers
  - Rebuild has the flavor of *ground* Knuth Bendix completion
  - Stock UF is uninterpreted values $e_i$ and atomic equations $e_i = e_j$
- Merges the concepts of containers, primitives, and e-ids
- E-nodes are interned, seids are ephemeral

# Decidable & Cheap

| seid | example | Canonizer |
|---|---|---|
| Atomic / Uninterp | $e_1$ | Union Find |
| primitive + uninterp | $Cons(7, e_1)$ | Value rooted UF + Unification |
| Group(oid) Action | $e_1 + 7$ | Group UF |
| Lin Expr | $2e_1 - 4e_7$ | Gauss Elim. / Row Echelon |
| Ground Terms | $foo(bar(e_7))$ | Inner E-Graph |

# Decidable & Expensive

| seid | example | Canonizer |
|---|---|---|
| Polynomials | $e_1 + 6e_4^3$ | Grobner Basis |
| Ground Multiset (AC) | $[e1, e1, e2]$ | Multiset KB / Graver / Hilbert bases |
| SMT Terms | | SMT sweeping |
| Bool Exprs | $e_1 \wedge e_2 \vee e_3$ | SAT Sweeping / BDDs / AIGs / Ordered Resolution |

## Strong (Undecidable) Theories

| seid | example | Canonizer |
|---|---|---|
| Strings (A) | $e_1 e_4 e_2$ | String Knuth Bendix |
| Terms w/ Vars | $foo(e_1, X)$ | Knuth Bendix |

# Wild Speculation

| seids | Example | Canonizer |
|---|---|---|
| Slotted eids? | $\lambda_{ijk} e_3(j, k, i)$ ? | ? |
| Colored eids? | $\Gamma \vdash e_{17}$ ? | ? |
| Non commutative Rings | $\partial_x e_1$ | ? |
| Towers | `Poly<MS<GroupAct<int>>>` | ? |
| Slotted Multisets | $e_{ijk} e_{jk}$ | ? |

## Related Work

- Normalized Rewriting (Marche)
- Alt-Ergo AC matching
- Extract, Rewrite, and Assert (Koehler et al)
- Mix E-nodes and Containers
- Brute Force SMT E-Graph
- Pavel's Blog Posts

# Thank You

- There is still much to do!
- Pre-print https://arxiv.org/abs/2504.14340
- Prototype: https://www.kdrag.com
    - `from kdrag.solvers.egraph import EGraph`