

February 1, 2023

A Functional Language with Hypergraphs as First-Class Data

Waseda University

Graduate School of Fundamental Science and Engineering

Master's Defence

Jin SANO

[twitter@sano_jn](https://twitter.com/sano_jn)

Overview

Imperative programmings with **heaps and pointers** are **tedious and error-prone**. E.g., Rust, C++,

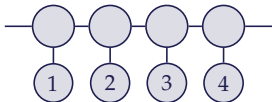


We construct a new **purely functional language** λ_{GT} , which handles **hypergraphs** as **immutable, first-class data** with **pattern matchings** based on *Graph Transformation* and a new **type system** F_{GT} .

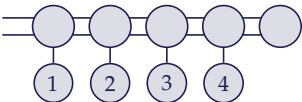
A Theoretical Foundation for
the Next Generation Programming Language.

Data Structures More Complex than Trees

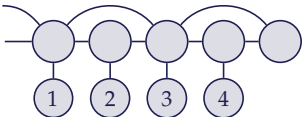
Difference List



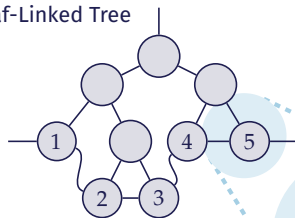
Doubly-Linked List



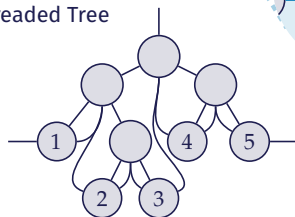
Skip List



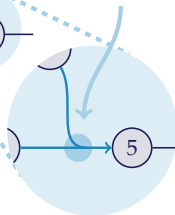
Leaf-Linked Tree



Threaded Tree



Hyperlink can
connect $n > 0$ vertices.



There are several important data structures that are beyond trees:
hypergraphs (= vertices + hyperlinks).

How Programming Paradigms Handle Data Structures

Imperative

- ! Heaps and pointers
- × Impure
- × Tedious and error-prone, and easily leads to significant security issues

Purely Functional

- ✓ Algebraic Data Types (trees)
- ✓ Pure
- ✓ Type system
- × Complex data structures are difficult to handle

Graph Transformations¹

- ✓ Graphs and pattern matchings on them
- × Impure

¹H. Ehrig et al. *Fundamentals of Algebraic Graph Transformation*. Monographs in Theoretical Computer Science. An EATCS Series. Springer, 2006.

Our Proposing Language λ_{GT} is ...

a functional language that extends
Algebraic Data Types (trees) to **hypergraphs**.

- ✓ Hypergraphs as first-class data
- ✓ Pattern matchings on hypergraphs
- ✓ Pure
- ✓ First-class functions
- ✓ Type system

Our Contributions

The main contributions of the thesis are threefold.

Chapter 2 We investigate the properties of the proposed hypergraph transformation formalism, HyperLMNtal².

Chapter 3 We propose λ_{GT} , a purely functional language that handles data structures beyond Algebraic Data Types, and implemented a PoC.

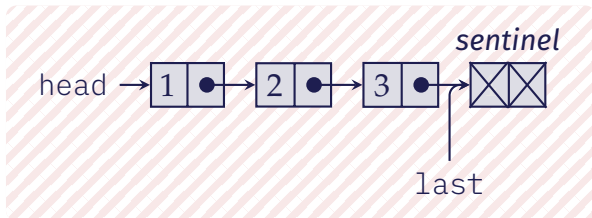
Chapter 4 We design a new type system F_{GT} for the λ_{GT} language.

Hereinafter, we may simply refer to hypergraphs as *graphs* and hyperlinks as *links*.

²J. Sano et al. “Syntax-driven and compositional syntax and semantics of Hypergraph Transformation System”. In: *Proc. 38th JSSST Annual Conference*. 2021.

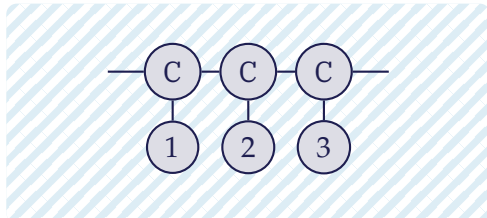
Example: Queues with Lists

Imperative



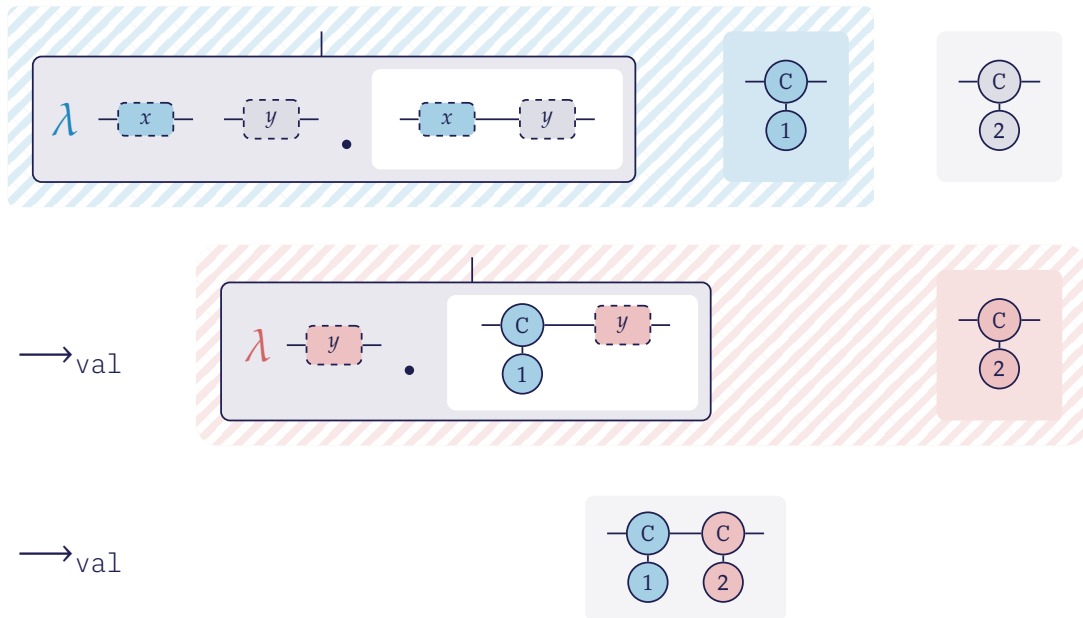
- ✗ Requires a **dummy** sentinel node.
- ✗ Low-level, tedious operations without a guarantee of the shape.

λ_{GT}



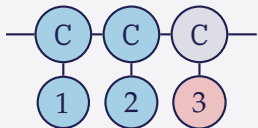
- ✓ Can define a *difference list*: a list with a link to the end.
- ✓ **Declarative operations.**
→ Next slides

Difference Lists Concatenation in λ_{GT}



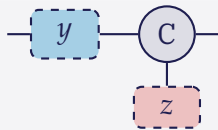
Pattern Matching Graphs

If  is bound with



, then

case  **of**



→

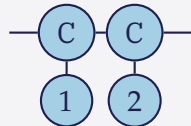


| otherwise

→



→_{val}



Pop the last element of a difference list

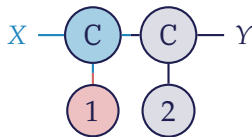
Chapter 2: Syntax of Graphs in HyperLMNtal

Graph

$G ::= \mathbf{0}$	Null	<i>empty graph</i>	$\vec{X} = X_1, \dots, X_n$
$p(\vec{X})$	Atom	<i>vertex with a name p and links \vec{X}</i>	
$X \bowtie Y$	Fusion	<i>link connection</i>	
(G, G)	Molecule	<i>composition of sub-graphs</i>	
$\nu X. G$	Hyperlink Creation	<i>scope of link names</i>	

For example, a *difference list* can be represented as

```
 $\nu Z. ($   
   $\nu Z_1. (\text{Cons}(Z_1, Z, X), 1(Z_1)),$   
   $\nu Z_2. (\text{Cons}(Z_2, Y, Z), 2(Z_2))$   
 $)$ 
```



Chapter 2: The Denoted Hypergraphs

We define the denoted hypergraphs and the mapping to them from HyperLMNtal terms, and prove that congruent terms are mapped to isomorphic graphs.

HyperLMNtal

$\nu Z.(\nu Z_1.(\text{Cons}(Z_1, Z, X), 1(Z_1)), \nu Z_2.(\text{Cons}(Z_2, Y, Z), 2(Z_2)))$

Equivalence is defined by
Structural Congruence

→

The Denoted Hypergraph

$\langle \{v_1, v_2\}, \langle \{X\}, \{\{X\}, \{Y\}\}, \{X \mapsto \{\langle v_1, 3 \rangle\}, Y \mapsto \{\langle v_2, 2 \rangle\}\}, \{\{\langle v_1, 1 \rangle, \langle v_2, 1 \rangle\}, \{\langle v_1, 2 \rangle, \langle v_3, 3 \rangle\}, \{\langle v_3, 1 \rangle, \langle v_4, 1 \rangle\}\}, \{v_1 \mapsto \text{Cons}, v_2 \mapsto 1, v_3 \mapsto \text{Cons}, v_4 \mapsto 2\} \rangle$

Equivalence is defined by
Graph Isomorphism

✓ Required for implementation justification and advanced verifications.

Chapter 3: Syntax of λ_{GT}

Value $G ::= \mathbf{0} \mid p(\vec{X}) \mid X \bowtie Y \mid (G, G) \mid \nu X. G$

Atom Name $p ::= C \mid \lambda x[\vec{X}]. e$

Expression $e ::= (\mathbf{case} \ e \ \mathbf{of} \ T \rightarrow e \mid \mathbf{otherwise} \rightarrow e) \mid (e \ e) \mid T$

Graph Template $T ::= \mathbf{0} \mid p(\vec{X}) \mid X \bowtie Y \mid (T, T) \mid \nu X. T \mid x[\vec{X}]$

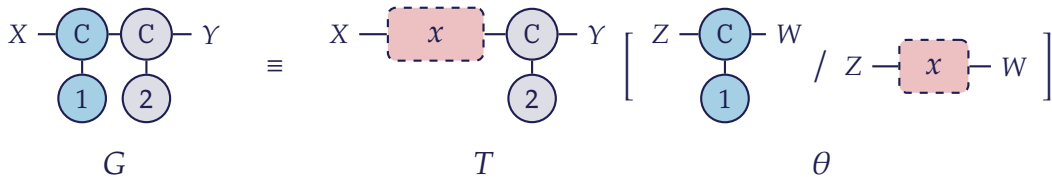
Wildcard

- ✓ Value in λ_{GT} is a **graph** in HyperLMNtal.
We allow *Constructor* and λ -*abstraction* for the atoms' names.
- ✓ λ_{GT} program is an **expression**.
- ✓ We use **Template = Value + Wildcards** for pattern matching.

Chapter 3: Pattern Matching with Graph Substitution

$$\frac{G \equiv T\vec{\theta}}{(\text{case } G \text{ of } T \rightarrow e_2 \mid \text{otherwise} \rightarrow e_3) \longrightarrow_{\text{val}} e_2\vec{\theta}} \text{ Rd-Case1}$$

For example,



Here, the graph G can be matched to the template T with a substitution θ .

Chapter 3: Reduction of λ_{GT}

$$\frac{G \equiv T\vec{\theta}}{(\text{case } G \text{ of } T \rightarrow e_2 \mid \text{otherwise} \rightarrow e_3) \longrightarrow_{\text{val}} e_2\vec{\theta}} \text{ Rd-Case1} \quad \text{Succeeded matching}$$

$$\frac{\neg \exists \vec{\theta}. (G \equiv T\vec{\theta})}{(\text{case } G \text{ of } T \rightarrow e_2 \mid \text{otherwise} \rightarrow e_3) \longrightarrow_{\text{val}} e_3} \text{ Rd-Case2} \quad \text{Failed matching}$$

$$\frac{fn(G) = \{\vec{X}\}}{((\lambda x[\vec{X}].e)(\vec{Y}) G) \longrightarrow_{\text{val}} e[G/x[\vec{X}]]} \text{ Rd-}\beta \quad \beta\text{-reduction}$$

$$\frac{e \longrightarrow_{\text{val}} e'}{E[e] \longrightarrow_{\text{val}} E[e']} \text{ Rd-Ctx} \quad \text{Call-by-value}$$

where $E ::= [] \mid (\text{case } E \text{ of } T \rightarrow e \mid \text{otherwise} \rightarrow e) \mid (E \ e) \mid (G \ E) \mid T$

Chapter 3: PoC Implementation

We build a reference implementation of the language
in only 500 lines of OCaml code.

Source <https://github.com/sano-jin/lambda-gt-alpha>
Try it at <https://sano-jin.github.io/lambda-gt-online>

We have run examples including ...

- ✓ Append two difference lists.
- ✓ Pop the last element of a difference list.
- ✓ Rotate a difference list (push an element to front from back).
- ✓ Pop all the elements from back of a difference list.
- ✓ Map a function on leaves of a leaf-linked tree.

Chapter 4: F_{GT} , a Type System for λ_{GT}

We propose a new type system, F_{GT} , for the λ_{GT} language.

The type in F_{GT} is a *type atom* $\tau(\vec{X})$:

- ✓ intuitively, a type in functional languages τ + links \vec{X} .

The typing relation $(\Gamma, P) \vdash e : \tau(\vec{X})$ denotes that

e has the type $\tau(\vec{X})$ under the type environment $\Gamma \stackrel{\text{def}}{=} \{x[\vec{X}] : \tau(\vec{X}), \dots\}$
and a set P of production rules, a graph grammar:

- ✓ an extension of a regular tree grammar,
on which Algebraic Data Types (trees) are based.

Chapter 4: Theorems of F_{GT}

We have proved some properties of F_{GT} .

Theorem 4.1 Soundness of F_{GT} .

Theorem 4.2 Correspondence between a typing relation in F_{GT} and a transitive closure of HyperLMNtal reduction.

- ✓ This allows us to take advantage of existing research of Graph Transformations³.

³P. Fradet et al. “Structured Gamma”. In: *Science of Computer Programming* 31.2 (1998), pp. 263–289; P. Fradet et al. “Shape types”. In: *Proc. POPL’97*. ACM. 1997, pp. 27–39; H. Björklund et al. “Uniform parsing for hyperedge replacement grammars”. In: *J. Computer and System Sciences* 118 (2021), pp. 1–27.

Related Work includes ...

FUnCAL⁴ is a functional language which supports graph-based database. They focus on a simple form of queries for the database, which is apart from our focus.

Structured Gamma⁵ and Shape Types⁶ provide a typing framework using graph grammar for graph transformation system applicable to verify imperative programs.

Separation Logic⁷ is a verification framework for imperative programs with heaps and pointers, which includes Cyclic Proof⁸ and SLRD⁹.

⁴K. Matsuda et al. “A Functional Reformulation of UnCAL Graph-Transformations: Or, Graph Transformation as Graph Reduction”. In: *Proc. POPL'97*. Paris, France: ACM, 2017, pp. 71–82.

⁵Fradet et al., “Structured Gamma”.

⁶Fradet et al., “Shape types”.

⁷J. Reynolds. “Separation logic: a logic for shared mutable data structures”. In: *Proc. LICS 2002*. IEEE. 2002, pp. 55–74.

⁸J. Brotherston et al. “A Generic Cyclic Theorem Prover”. In: *Proc. APLAS 2012*. Vol. 7705. Lecture Notes in Computer Science. Springer, 2012, pp. 350–367.

⁹R. Iosif et al. “The Tree Width of Separation Logic with Recursive Definitions”. In: *Automated Deduction – CADE-24*. 2013, pp. 21–38.

Summary and Future Work

We introduced ...

- ✓ **HyperLMNtal**: a hypergraph transformation formalism,
- ✓ λ_{GT} : a new functional language with hypergraphs as first-class data, and
- ✓ F_{GT} : a new type system for the λ_{GT} language.

Artifact: <https://github.com/sano-jin/lambda-gt-alpha>

Future work includes ...

- ✓ more investigations on HyperLMNtal properties,
- ✓ the justification of the PoC, the construction of an efficient compiler, and
- ✓ more advanced verifications on F_{GT} .

Publications

1. J. Sano et al. “Syntax-driven and compositional syntax and semantics of Hypergraph Transformation System”. In: *Proc. 38th JSSST Annual Conference*. 2021. Unrefereed. **Student encouragement award.**
2. J. Sano et al. “A functional language with graphs as first-class data”. In: *Proc. 39th JSSST Annual Conference*. 2022. Unrefereed. **Presentation award.**
3. J. Sano et al. “Type Checking Data Structures More Complex Than Trees”. In: *Journal of Information Processing and IPSJ Transactions on Programming* (2023). Accepted.
4. J. Sano et al. “Axiomatizing Hypergraph Isomorphism”. In: *Special Interest Group on Programming and Programming Language*. 2023. Accepted.

Appendix

Related Work

HyperLMNtal

Lambda GT

FGT

Related Implementations

	Implemented with ...	LOC
λ_{GT} Reference Interpreter	OCaml	500
GP 2 Reference Interpreter ¹⁰	Haskell	1,000
HyperLMNtal Compiler/Runtime ¹¹	Java/C++	12,000/47,000

¹⁰C. Bak et al. “A Reference Interpreter for the Graph Programming Language GP 2”. In: *Proceedings Graphs as Models, GaM@ETAPS 2015, London, UK, 11-12 April 2015*. Ed. by A. Rensink et al. Vol. 181. EPTCS. 2015, pp. 48–64.

¹¹LMNtal. <https://github.com/lmntal/lmntal-compiler>. (Visited on 08/10/2022); SLIM. <https://github.com/lmntal/slim>. (Visited on 08/10/2022); M. Gocho et al. “Evolution of the LMNtal Runtime to a Parallel Model Checker”. In: *Computer Software* 28.4 (2011), 4_137–4_157.

Appendix

Related Work

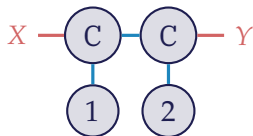
HyperLMNtal

Lambda GT

FGT

Free Names and Substitutions of Hyperlinks

Links bound by ν are called *Local Links* and others are called *Free Links*

$$\begin{aligned} &\nu Z_1.(\\ &\quad \nu Z_2.(\text{Cons}(Z_1, Z, X), 1(Z_1)), \\ &\quad \nu Z_2.(\text{Cons}(Z_2, Y, Z), 2(Z_2)) \\ &)\end{aligned}$$


- $fn(G)$ denotes the set of all free links in G
- $G\langle \vec{Y}/\vec{X} \rangle$ replaces all free occurrences of \vec{X} with \vec{Y} .

The notion of locality of (link) names is NOT common in graph formalisms but in the formalisms for PLs; λ -calculus, π -calculus, ...

Structural Congruence: Axioms of Graph Equivalences

- (E1) $(\mathbf{0}, G) \equiv G$
(E2) $(G_1, G_2) \equiv (G_2, G_1)$
(E3) $(G_1, (G_2, G_3)) \equiv ((G_1, G_2), G_3)$
(E4) $G_1 \equiv G_2 \Rightarrow (G_1, G_3) \equiv (G_2, G_3)$
(E5) $G_1 \equiv G_2 \Rightarrow \nu X. G_1 \equiv \nu X. G_2$
(E6) $\nu X. (X \bowtie Y, G) \equiv \nu X. G \langle Y/X \rangle$
where $X \in \text{fn}(G) \vee Y \in \text{fn}(G)$
(E7) $\nu X. \nu Y. X \bowtie Y \equiv \mathbf{0}$
(E8) $\nu X. \mathbf{0} \equiv \mathbf{0}$
(E9) $\nu X. \nu Y. G \equiv \nu Y. \nu X. G$
(E10) $\nu X. (G_1, G_2) \equiv (\nu X. G_1, G_2)$
where $X \notin \text{fn}(G_2)$

For example,

$$\begin{aligned} & \nu Z. (\\ & \quad \nu Z_1. (\text{Cons}(Z_1, Z, X), 1(Z_1)), \\ & \quad \nu Z_2. (\text{Cons}(Z_2, Y, Z), 2(Z_2)) \\ &) \\ & \equiv \\ & \nu Z. (\\ & \quad \nu Z_1. (1(Z_1), \text{Cons}(Z_1, Z, X)), \\ & \quad \nu Z_2. (\text{Cons}(Z_2, Y, Z), 2(Z_2)) \\ &) \\ & \text{by (E2), (E4) and (E5)} \end{aligned}$$

✓ Notice the rules are defined compositionally.

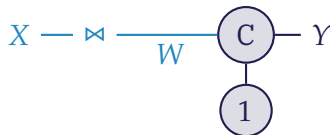
Fusion

Structural Congruence

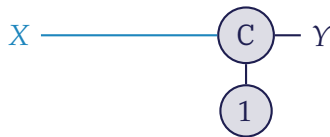
$$(E6) \quad \nu X.(X \bowtie Y, G) \equiv \nu X.G\langle Y/X \rangle$$

where $X \in fn(G) \vee Y \in fn(G)$

$$\nu WZ.(W \bowtie X, \text{Cons}(Z, Y, W), 1(Z))$$



$$\equiv \nu WZ.(\text{Cons}(Z, Y, X), 1(Z))$$



Appendix

Related Work

HyperLMNtal

Lambda GT

FGT

Graph Template

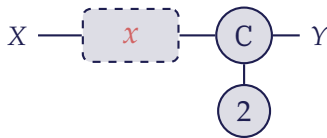
Graph Template

$T ::= \mathbf{0} \mid p(\vec{X}) \mid X \bowtie Y \mid (T, T) \mid vX.T$
 $\mid x[\vec{X}]$ **Graph context** *wildcard in pattern matching; variable*

Since the value in λ_{GT} is Graph, we use *Template* of graphs to represent data with variables.

For example,

$vZ.($
 $x[Z, X],$
 $vZ_2.(\text{Cons}(Z_2, Y, Z), 2(Z_2))$
 $)$



Graph Substitution

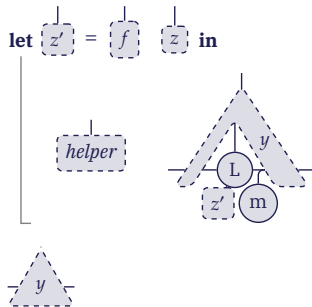
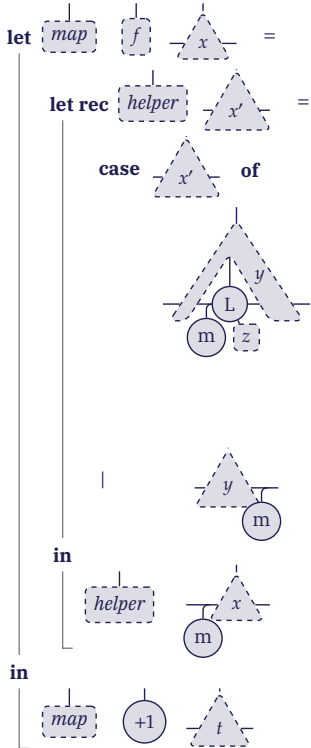
We define capture-avoiding substitution θ of a graph context $x[\vec{X}]$ with a template T in e , written $e[T/x[\vec{X}]]$.

- The definition is standard except for the graph contexts.

$$\begin{aligned} (x[\vec{X}])[T/y[\vec{Y}]] &= \\ \text{if } x/|\vec{X}| &= y/|\vec{Y}| \text{ then } T\langle\vec{X}/\vec{Y}\rangle && \text{reconnect free links} \\ \text{else } x[\vec{X}] & \end{aligned}$$

For example,





Map a function on leaves of a leaf-linked tree. 30/19

Appendix

Related Work

HyperLMNtal

Lambda GT

FGT

Chapter 4: Syntax of F_{GT}

The type in F_{GT} is a type atom $\tau(\vec{X})$ where ...

Atom Name for Types $\tau ::= \alpha \mid \tau(\vec{X}) \rightarrow \tau(\vec{X})$

Production Rule $r ::= \alpha(\vec{X}) \rightarrow \mathcal{T}$

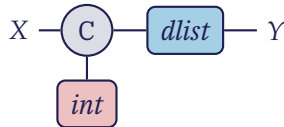
Type Graph $\mathcal{T} ::= \tau(\vec{X}) \mid C(\vec{X}) \mid X \bowtie Y \mid (\mathcal{T}, \mathcal{T}) \mid \nu X. \mathcal{T}$

For example, the production rules of difference lists are:

$X \text{---} \boxed{dlist} \text{---} Y \quad dlist(Y, X) \rightarrow X \bowtie Y$

$\mid \nu Z_1. \nu Z_2. ((\text{Cons}(Z_1, Z_2, X),$
 $\quad \text{int}(Z_1)), dlist(Y, Z_2))$

$X \text{-----} Y$



Rules of F_{GT} $\langle 1/2 \rangle$: Typing Rules as in Functional Languages

- These are basically the same as the type system of the other ordinary functional languages, except that **the type in F_{GT} is an atom**.

$$\frac{(\Gamma, P) \vdash e_1 : (\tau_1(\vec{X}) \rightarrow \tau_2(\vec{Y}))(\vec{Z}) \quad (\Gamma, P) \vdash e_2 : \tau_1(\vec{X})}{(\Gamma, P) \vdash (e_1 \ e_2) : \tau_2(\vec{Y})} \text{Ty-App}$$

$$\frac{((\Gamma, x[\vec{X}] : \tau_1(\vec{X})), P) \vdash e : \tau_2(\vec{Z})}{(\Gamma, P) \vdash (\lambda x[\vec{X}] : \tau_1(\vec{Y}).e)(\vec{W}) : (\tau_1(\vec{Y}) \rightarrow \tau_2(\vec{Z}))(\vec{W})} \text{Ty-Arrow}$$

$$\frac{}{(\Gamma\{x[\vec{X}] : \tau(\vec{Y})\}, P) \vdash x[\vec{X}] : \tau(\vec{Y})} \text{Ty-Var}$$

$$\frac{(\Gamma, P) \vdash e_1 : \tau_1(\vec{X}) \quad ((\Gamma, \Gamma'), P) \vdash e_2 : \tau_2(\vec{Y}) \quad (\Gamma, P) \vdash e_3 : \tau_2(\vec{Y})}{(\Gamma, P) \vdash (\text{case } e_1 \text{ of } T \rightarrow e_2 \mid \text{otherwise} \rightarrow e_3) : \tau_2(\vec{Y})} \text{Ty-Case}$$

* We gave a detailed explanation of Γ' in the paper.

Rules of F_{GT} <2/2>: Typing Rules for Graphs

$$\frac{(\Gamma, P) \vdash T : \tau(\vec{X}) \quad T \equiv T'}{(\Gamma, P) \vdash T' : \tau(\vec{X})} \text{Ty-Cong}$$

$$\frac{(\Gamma, P) \vdash T : \tau(\vec{X})}{(\Gamma, P) \vdash T\langle Z/Y \rangle : \tau(\vec{X})\langle Z/Y \rangle} \text{Ty-Alpha}$$

where $Z \notin \text{fn}(T)$

$$\frac{(\Gamma, P) \vdash T_1 : \tau_1(\vec{X}_1) \quad \dots \quad (\Gamma, P) \vdash T_n : \tau_n(\vec{X}_n)}{(\Gamma, P\{\alpha(\vec{X}) \longrightarrow \mathcal{T}\}) \vdash \mathcal{T}[T_1/\tau_1(\vec{X}_1), \dots, T_n/\tau_n(\vec{X}_n)] : \alpha(\vec{X})} \text{Ty-Prod}$$

where $\tau_i(\vec{X}_i)$ are all the type variable or arrow atoms appearing in \mathcal{T}

Ty-Prod Example

$$\frac{(\Gamma, P) \vdash T_1 : \tau_1(\vec{X}_1) \quad \dots \quad (\Gamma, P) \vdash T_n : \tau_n(\vec{X}_n)}{(\Gamma, P\{\alpha(\vec{X}) \longrightarrow \mathcal{T}\}) \vdash \mathcal{T} [T_1/\tau_1(\vec{X}_1), \dots, T_n/\tau_n(\vec{X}_n)] : \alpha(\vec{X})} \text{Ty-Prod}$$

where $\tau_i(\vec{X}_i)$ are all the type variable or arrow atoms appearing in \mathcal{T}

For example, for

$$nodes(Y, X) \longrightarrow \nu Z_1. \nu Z_2. (\text{Cons}(Z_1, Z_2, X), \text{nat}(Z_1), nodes(Y, Z_2)) \quad \dots \quad r_2$$

the Ty-Prod is

$$\frac{(\Gamma, P) \vdash T_1 : \text{nat}(Z_1) \quad (\Gamma, P) \vdash T_2 : nodes(Y, Z_2)}{(\Gamma, P\{P_2\}) \vdash \nu Z_1. \nu Z_2. (\text{Cons}(Z_1, Z_2, X), \text{nat}(Z_1), nodes(Y, Z_2)) [T_1/\text{nat}(Z_1), T_2/nodes(Y, Z_2)]} \text{Ty-Prod}$$

$$= \nu Z_1. \nu Z_2. (\text{Cons}(Z_1, Z_2, X), T_1, T_2) : nodes(Y, X)$$

Example: Typing a Difference List

$$(\{n[Z_1] : \text{nat}(Z_1)\}, \{r_1, r_2\}) \vdash \text{Cons}(n, Y, X) : \text{nodes}(Y, X)$$

where r_1 and r_2 are the followings.

$$\text{nodes}(Y, X) \longrightarrow X \bowtie Y$$

$$\text{nodes}(Y, X) \longrightarrow \text{Cons}(\text{nat}, \text{nodes}(Y), X)$$

can be shown as follows.

$$\frac{\frac{}{(\Gamma, P) \vdash n[Z_1] : \text{n}(Z_1)} \text{Ty-Var} \quad \frac{\frac{}{(\Gamma, P\{r_1\}) \vdash X \bowtie Y : \text{nodes}(Y, X)} \text{Ty-Prod} \quad \frac{}{(\Gamma, P) \vdash Z_2 \bowtie Y : \text{nodes}(Z_2, X)} \text{Ty-Alpha}}{(\Gamma, P\{r_2\}) \vdash T' : \text{nodes}(Y, X) \quad \text{where} \quad T' = \nu Z_1 Z_2. (\text{Cons}(Z_1, Z_2, X), n[Z_1], Z_2 \bowtie Y)} \text{Ty-Prod} \quad T \equiv T'}{(\Gamma, P) \vdash T : \text{nodes}(Y, X) \quad \text{where} \quad T = \text{Cons}(\text{succ}, Y, X)} \text{Ty-Cong}$$