

# FROM SYSTEM F TO TYPED ASSEMBLY LANGUAGE

---

*Greg Morrisett, David Walker, Karl Crary & Neal Glew  
TOPLAS 1999*

*Presentation by:  
Drew Zagieboylo/Matthew Milano*

# TYPED ASSEMBLY LANGUAGE

---

|                             |   |
|-----------------------------|---|
| <i>types</i>                | $\tau, \sigma ::= \alpha \mid int \mid \forall[\vec{\alpha}].\Gamma \mid \langle \tau_1^{\varphi_1}, \dots, \tau_n^{\varphi_n} \rangle \mid \exists\alpha.\tau$ |
| <i>initialization flags</i> | $\varphi ::= 0 \mid 1$  |
| <i>heap types</i>           | $\Psi ::= \{l_1:\tau_1, \dots, l_n:\tau_n\}$  |
| <i>register file types</i>  | $\Gamma ::= \{r_1:\tau_1, \dots, r_n:\tau_n\}$  |
| <i>type contexts</i>        | $\Delta ::= \alpha_1, \dots, \alpha_n$  |

# TYPED ASSEMBLY LANGUAGE

---

*types*  $\tau, \sigma ::= \alpha \mid \text{int} \mid \forall[\vec{\alpha}].\Gamma \mid \langle \tau_1^{\varphi_1}, \dots, \tau_n^{\varphi_n} \rangle \mid \exists\alpha.\tau$   
*initialization flags*  $\varphi ::= 0 \mid 1$   
*heap types*  $\Psi ::= \{\ell_1:\tau_1, \dots, \ell_n:\tau_n\}$   
*register file types*  $\Gamma ::= \{r_1:\tau_1, \dots, r_n:\tau_n\}$   
*type contexts*  $\Delta ::= \alpha_1, \dots, \alpha_n$

*registers*  $r ::= \mathbf{r1} \mid \mathbf{r2} \mid \mathbf{r3} \mid \dots$   
*word values*  $w ::= \ell \mid i \mid ?\tau \mid w[\tau] \mid \mathbf{pack}[\tau, w] \text{ as } \tau'$   
*small values*  $v ::= r \mid w \mid v[\tau] \mid \mathbf{pack}[\tau, v] \text{ as } \tau'$   
*heap values*  $h ::= \langle w_1, \dots, w_n \rangle \mid \mathbf{code}[\vec{\alpha}]\Gamma.I$   
*heaps*  $H ::= \{\ell_1 \mapsto h_1, \dots, \ell_n \mapsto h_n\}$   
*register files*  $R ::= \{r_1 \mapsto w_1, \dots, r_n \mapsto w_n\}$

*instructions*  $\iota ::= \mathbf{add} r_d, r_s, v \mid \mathbf{bnz} r, v \mid \mathbf{ld} r_d, r_s[i] \mid \mathbf{malloc} r_d[\vec{\tau}] \mid \mathbf{mov} r_d, v \mid$   
 $\mathbf{mul} r_d, r_s, v \mid \mathbf{st} r_d[i], r_s \mid \mathbf{sub} r_d, r_s, v \mid \mathbf{unpack}[\alpha, r_d], v$   
*instruction sequences*  $I ::= \iota; I \mid \mathbf{jmp} v \mid \mathbf{halt}[\tau]$   
*programs*  $P ::= (H, R, I)$

# TYPED ASSEMBLY LANGUAGE

---

*types*  $\tau, \sigma ::= \alpha \mid \text{int} \mid \forall[\vec{\alpha}].\Gamma \mid \langle \tau_1^{\varphi_1}, \dots, \tau_n^{\varphi_n} \rangle \mid \exists\alpha.\tau$   
*initialization flags*  $\varphi ::= 0 \mid 1$   
*heap types*  $\Psi ::= \{\ell_1:\tau_1, \dots, \ell_n:\tau_n\}$   
*register file types*  $\Gamma ::= \{r_1:\tau_1, \dots, r_n:\tau_n\}$   
*type contexts*  $\Delta ::= \alpha_1, \dots, \alpha_n$

*registers*  $r ::= \mathbf{r1} \mid \mathbf{r2} \mid \mathbf{r3} \mid \dots$   
*word values*  $w ::= \ell \mid i \mid ?\tau \mid w[\tau] \mid \text{pack}[\tau, w] \text{ as } \tau'$   
*small values*  $v ::= r \mid w \mid v[\tau] \mid \text{pack}[\tau, v] \text{ as } \tau'$   
*heap values*  $h ::= \langle w_1, \dots, w_n \rangle \mid \text{code}[\vec{\alpha}]\Gamma.I$   
*heaps*  $H ::= \{\ell_1 \mapsto h_1, \dots, \ell_n \mapsto h_n\}$   
*register files*  $R ::= \{r_1 \mapsto w_1, \dots, r_n \mapsto w_n\}$

*instructions*  $\iota ::= \text{add } r_d, r_s, v \mid \text{bnz } r, v \mid \text{ld } r_d, r_s[i] \mid \text{malloc } r_d[\vec{\tau}] \mid \text{mov } r_d, v \mid$   
 $\text{mul } r_d, r_s, v \mid \text{st } r_d[i], r_s \mid \text{sub } r_d, r_s, v \mid \text{unpack}[\alpha, r_d], v$   
*instruction sequences*  $I ::= \iota; I \mid \text{jmp } v \mid \text{halt}[\tau]$   
*programs*  $P ::= (H, R, I)$



**WHY DO WE WANT TAL?**

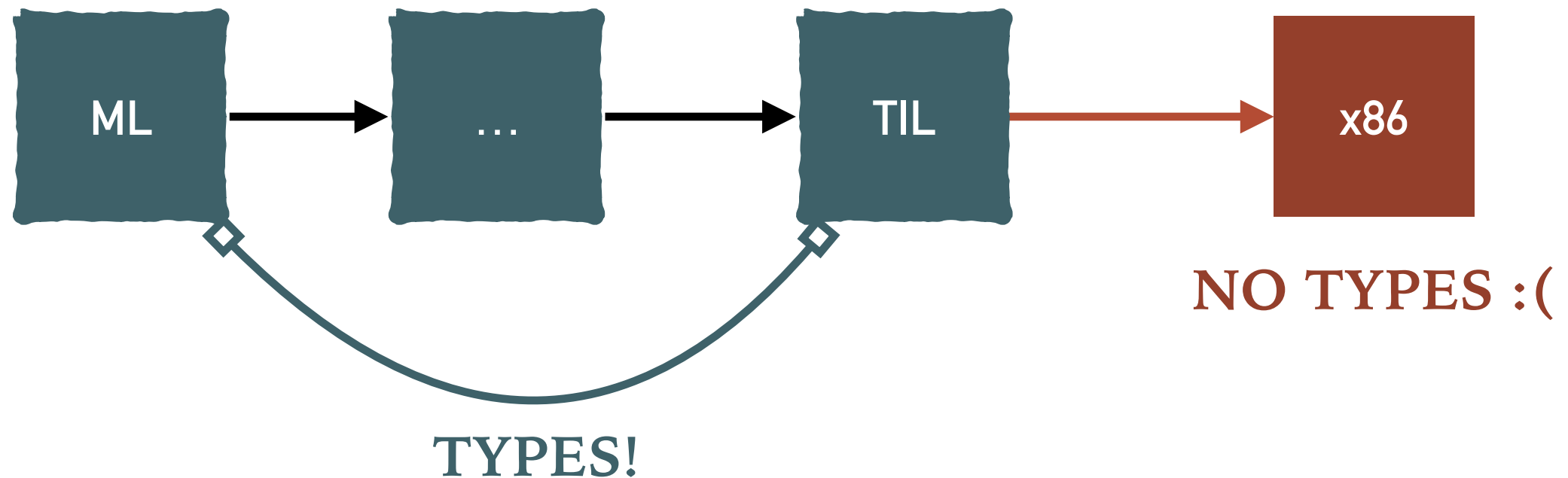
**TYPE SYSTEMS ALL THE  
WAY!!**

# TYPED INTERMEDIATE LANGUAGES

---

## ➤ TIL

- Throughout the 90's (and today!)
- Benefits of Types (efficiency + soundness)
- Target Language is *Untyped*



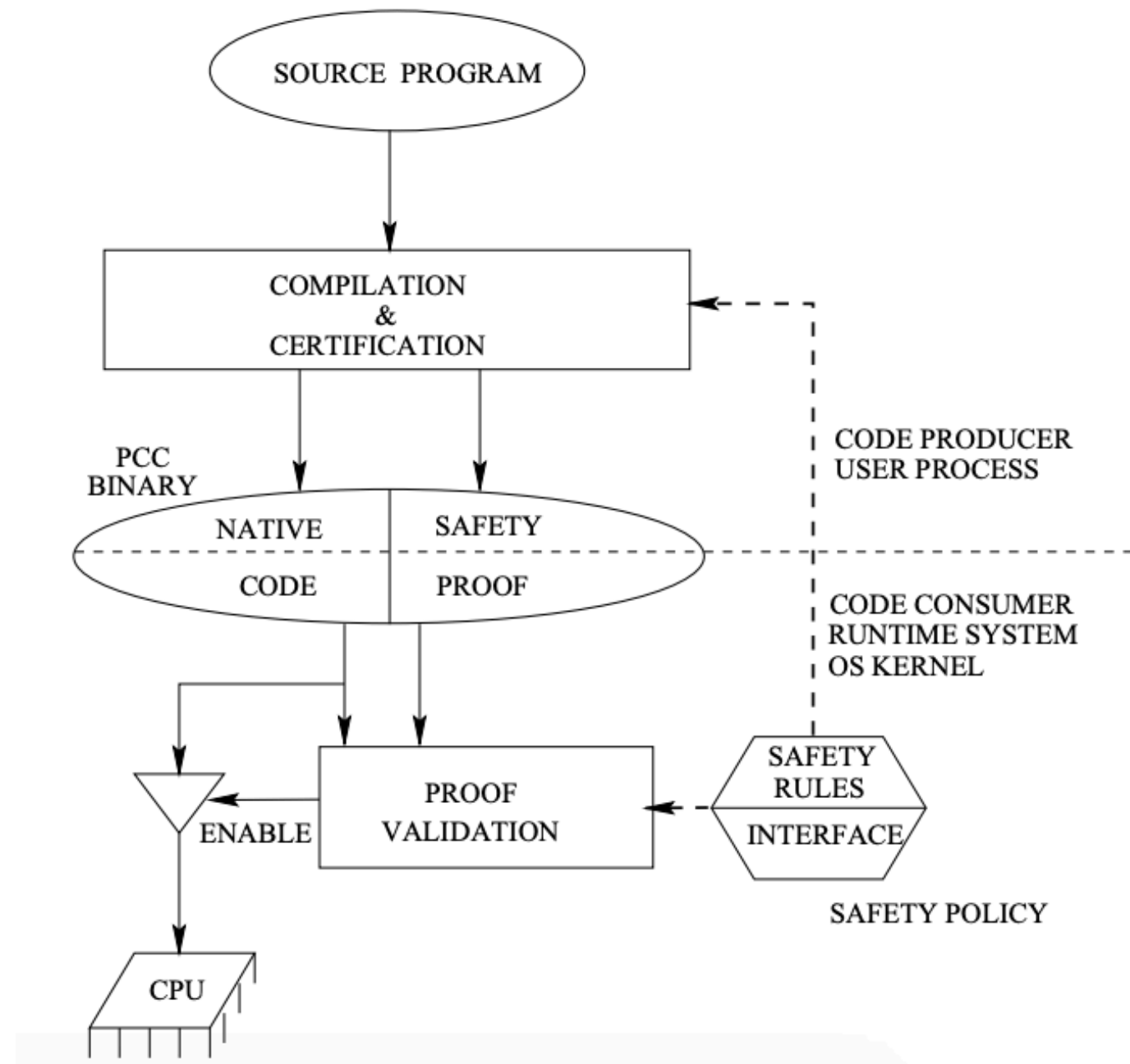


**HOW TO GUARANTEE  
SAFETY W/ UNTYPED  
AND UNTRUSTED CODE?**

# PROOF-CARRYING CODE

---

- George Necula (POPL '97)
- Compiler Produces:
  1. Program
  2. Proof
- First-Order Predicate Logic Based
- Difficult to Build Compilers





# TYPED ASSEMBLY LANGUAGE – FEATURES

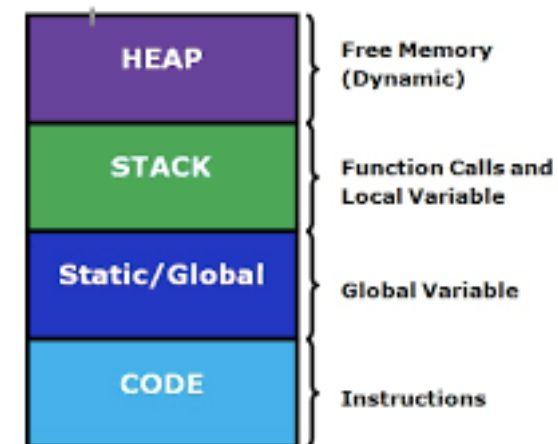
---

- RISC-style language
- Types:
  - Code types
  - Pointer Types
  - Existential Type Constructor
- Security:
  - No pointer forging!
  - Control Flow Integrity
- Other:
  - Memory Allocation

```
l_fact:  
code[] {r1:int}.
```



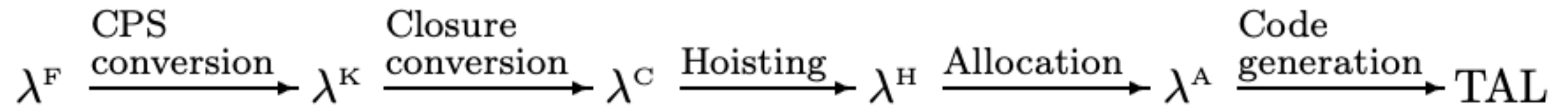
Application Memory



# SYSTEM F TO TAL

---

- Show that TAL is *expressive*



# SYSTEM F TO TAL

---

## ➤ CPS Conversion



# CPS TRANSLATION

---

- Continuation Passing Style
  - Translate to near-linear series of let bindings & calls
  - Removes function call stack

## *Abstraction Translation*

$$\mathcal{K}_{\text{exp}}\llbracket(\text{fix } x(x_1:\tau_1):\tau_2.e)^\tau\rrbracket k \stackrel{\text{def}}{=} k((\text{fix } x(x_1:\mathcal{K}\llbracket\tau_1\rrbracket), c:\mathcal{K}_{\text{cont}}\llbracket\tau_2\rrbracket).\mathcal{K}_{\text{exp}}\llbracket e\rrbracket c^{\mathcal{K}_{\text{cont}}\llbracket\tau_2\rrbracket})^{\mathcal{K}\llbracket\tau\rrbracket})$$

## *Application Translation*

$$\mathcal{K}_{\text{exp}}\llbracket(u_1^{\tau_1} u_2^{\tau_2})^\tau\rrbracket k \stackrel{\text{def}}{=} \mathcal{K}_{\text{exp}}\llbracket u_1^{\tau_1}\rrbracket(\lambda x_1:\mathcal{K}\llbracket\tau_1\rrbracket.\mathcal{K}_{\text{exp}}\llbracket u_2^{\tau_2}\rrbracket(\lambda x_2:\mathcal{K}\llbracket\tau_2\rrbracket.x_1^{\mathcal{K}\llbracket\tau_1\rrbracket}(x_2^{\mathcal{K}\llbracket\tau_2\rrbracket}, k))^{\mathcal{K}_{\text{cont}}\llbracket\tau_2\rrbracket})^{\mathcal{K}_{\text{cont}}\llbracket\tau_1\rrbracket})$$

# SYSTEM F TO $\lambda_K$

---

## ► Continuation Passing Style

$\lambda_F$   $(\text{fix } f(n : \text{int}) : \text{int} . \text{if0 } (n, 1, n \times f(n - 1))) 6$



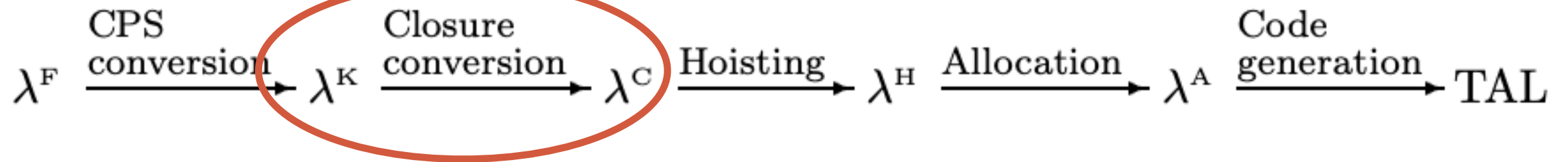
$\lambda_K$   $(\text{fix } f(n : \text{int}, k : (\text{int}) \rightarrow \text{void}) .$   
     $\text{if0}(n, k(1),$   
         $\text{let } x = n - 1 \text{ in}$   
         $f(x, \lambda(y : \text{int}) . \text{let } z = n \times y \text{ in } k(z))))$   
 $(6, \lambda(n : \text{int}) . \text{halt}[\text{int}]n)$



# SYSTEM F TO TAL

---

## ► Closure Conversion



# POLYMORPHIC CLOSURE CONVERSION

---

- Generate Explicit Closures
- Implements Encapsulation
- New Syntax

- Existential Types

$$\tau, \sigma ::= \dots \mid \exists \alpha . \tau$$

- Packing/Unpacking

$$u ::= \dots \mid v[\tau] \mid \text{pack}[\tau_1, v] \text{ as } \tau_2$$

$$d ::= \dots \mid [\alpha, x] = \text{unpack } v$$

- Uses *Type Erasure*\*
- Function bodies type-check w/o environment type info
- **Pack** is a no-op at runtime

# $\lambda_K$ TO $\lambda_C$

---

## ► Polymorphic Closure Conversion

### *Function Type Translation*

$$\mathcal{C}[\forall[\vec{\alpha}].(\tau_1, \dots, \tau_n) \rightarrow \text{void}] = \exists\beta. \langle \forall[\vec{\alpha}].(\beta, \mathcal{C}[\tau_1], \dots, \mathcal{C}[\tau_n]) \rightarrow \text{void}, \beta \rangle$$

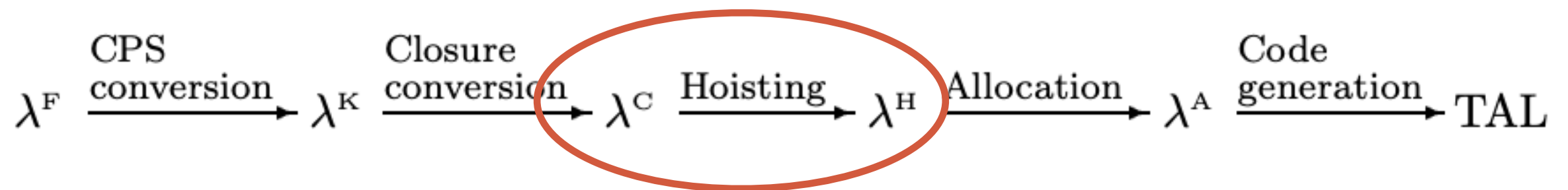
### *Application Translation*

$$\begin{aligned} \mathcal{C}_{\text{exp}}[u^\tau[\sigma_1, \dots, \sigma_m](v_1, \dots, v_n)] &\stackrel{\text{def}}{=} \text{let } [\gamma, z] = \text{unpack } \mathcal{C}_{\text{val}}[u^\tau] \text{ in} \\ &\text{let } z_{\text{code}} = \pi_1(z^{\langle \tau_{\text{code}}, \gamma \rangle}) \text{ in} \\ &\text{let } z_{\text{env}} = \pi_2(z^{\langle \tau_{\text{code}}, \gamma \rangle}) \text{ in} \\ &(z_{\text{code}}^{\tau_{\text{code}}}[\mathcal{C}[\sigma_1], \dots, \mathcal{C}[\sigma_m]]) \\ &\quad (z_{\text{env}}^\gamma, \mathcal{C}_{\text{val}}[v_1], \dots, \mathcal{C}_{\text{val}}[v_n]) \\ &\text{where} \\ &\mathcal{C}[\tau] = \exists\gamma. \langle \tau_{\text{code}}, \gamma \rangle \end{aligned}$$

# SYSTEM F TO TAL

---

## ► Hoisting



# HOISTING

---

- Separating Code Definition & Program
- Much like real memory layout
  - Closures make this easy!
  - Bind `fix` statements to variables, pointing to code

Syntax changes:

|                    |  |
|--------------------|--|
| <i>values</i>      | $u ::= \text{delete fix } x(x_1:\tau_1, \dots, x_n:\tau_n).e$                |
| <i>heap values</i> | $h ::= \text{code}[\vec{\alpha}](x_1:\tau_1, \dots, x_n:\tau_n).e$           |
| <i>programs</i>    | $P ::= \text{letrec } x_1 \mapsto h_1, \dots, x_n \mapsto h_n \text{ in } e$ |

The typing rule for `fix` is replaced by a heap value rule for `code`:

$$\frac{\vec{\alpha} \vdash_{\text{H}} \tau_i \quad \vec{\alpha}; (\Gamma, x_1:\tau_1, \dots, x_n:\tau_n) \vdash_{\text{H}} e}{\Gamma \vdash_{\text{H}} \text{code}[\vec{\alpha}](x_1:\tau_1, \dots, x_n:\tau_n).e : \forall[\vec{\alpha}](\tau_1, \dots, \tau_n) \rightarrow \text{void} \text{ hval}} \quad (x_1, \dots, x_n \notin \Gamma)$$

New typing rule:

$$\frac{\emptyset \vdash_{\text{H}} \tau_i \quad x_1:\tau_1, \dots, x_n:\tau_n \vdash_{\text{H}} h_i : \tau_i \text{ hval} \quad \emptyset; x_1:\tau_1, \dots, x_n:\tau_n \vdash_{\text{H}} e}{\vdash_{\text{H}} \text{letrec } x_1 \mapsto h_1, \dots, x_n \mapsto h_n \text{ in } e} \quad (x_i \neq x_j \text{ for } i \neq j)$$

$\lambda_K$   
 .....  
 ➤ P  
 ➤ F

```

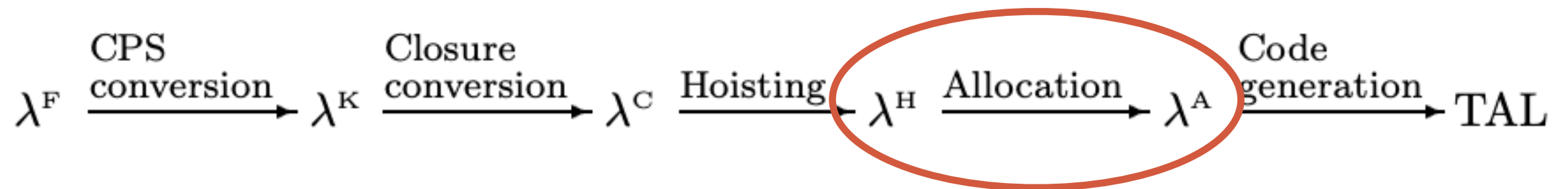
letrec f_code   ↦ (* main factorial code block *)
  code[](env:⟨⟩, n:int, k:τk).
    if0(n, (* true branch: continue with 1 *)
      let [β, k_unpack] = unpack k in
      let k_code = π1(k_unpack) in
      let k_env = π2(k_unpack) in
      k_code(k_env, 1),
      (* false branch: recurse with n - 1 *)
      let x = n - 1 in
      f_code(env, x, pack [⟨int, τk⟩, ⟨cont_code, ⟨n, k⟩⟩] as τk))
cont_code ↦ (* code block for continuation after factorial computation *)
  code[](env:⟨int, τk⟩, y:int).
    (* open the environment *)
    let n = π1(env) in
    let k = π2(env) in
    (* compute n! into z *)
    let z = n × y in
    (* continue with z *)
    let [β, k_unpack] = unpack k in
    let k_code = π1(k_unpack) in
    let k_env = π2(k_unpack) in
    k_code(k_env, z)
halt_code ↦ (* code block for top-level continuation *)
  code[](env:⟨⟩, n:int). halt[int]n
in
  f_code(⟨⟩, 6, pack [⟨⟩, ⟨halt_code, ⟨⟩⟩] as τk)
  
```

where  $\tau_k$  is  $\exists\alpha.\langle(\alpha, \text{int}) \rightarrow \text{void}, \alpha\rangle$

# SYSTEM F TO TAL

---

## ► Memory Allocation



# ALLOCATION

---

- ▶ Assembly language doesn't have Tuples!
- ▶ Need to allocate memory for tuples (and initialize!)

$$A[[\langle \tau_1, \dots, \tau_n \rangle]] \triangleq \langle A[[\tau_1]]^1, \dots, A[[\tau_n]]^1 \rangle$$

- ▶  $x = (v_1, v_2)$

let  $x_1 : \langle int^0, int^0 \rangle = \text{malloc}[int, int]$

$x_2 : \langle int^1, int^0 \rangle = x_1[1] \leftarrow v_1$

$x : \langle int^1, int^1 \rangle = x_2[2] \leftarrow v_2$

⋮



# ALLOCATION

---

$\lambda_H$

```
(* false branch: recurse with  $n - 1$  *)  
let  $x = n - 1$  in  
 $f_{code}(env, x, \text{pack} [\langle int, \tau_k \rangle, \langle cont_{code}, \langle n, k \rangle \rangle] \text{ as } \tau_k))$ 
```



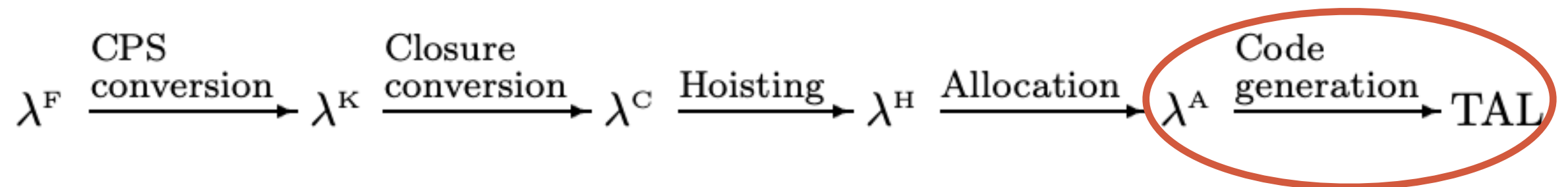
$\lambda_A$

```
let  $x = n - 1$  in  
let  $y_1 = \text{malloc}[int, \tau_k]$  in  
let  $y_2 = y_1[1] \leftarrow n$  in  
let  $y_3 = y_2[2] \leftarrow k$  in      (*  $\langle n, k \rangle$  *)  
let  $y_4 = \text{malloc}[(\langle int, \tau_k \rangle, int) \rightarrow void, \langle int, \tau_k \rangle]$  in  
let  $y_5 = y_4[1] \leftarrow cont_{code}$  in  
let  $y_6 = y_5[2] \leftarrow y_3$  in      (*  $\langle cont_{code}, \langle n, k \rangle \rangle$  *)  
 $f_{code}(env, x, \text{pack} [\langle int, \tau_k \rangle, y_6] \text{ as } \tau_k))$ 
```

# SYSTEM F TO TAL

---

## ► Code Generation



# SYSTEM F TO TAL

---

- Code Generation
  - Mostly direct translation to assembly
  - Function types annotate registers

$$\mathcal{T}[\forall[\vec{\alpha}](\tau_1, \dots, \tau_n) \rightarrow \text{void}] \stackrel{\text{def}}{=} \forall[\vec{\alpha}]\{\mathbf{r}1:\mathcal{T}[\tau_1], \dots, \mathbf{r}n:\mathcal{T}[\tau_n]\}$$

- **unpack** is just a **mov** instruction w/ type erasure
- **malloc** is abstract

# TAL IMPLEMENTATION

---

- TALx86 : IA32 ISA
- Variation from Paper:
  - Other data types (arrays, floats, etc.)
  - Not CPS -> Uses Explicit Stack
  - Implements **malloc** and **unpack** instructions
  - Modules with Type Interfaces
- Some optimizations
  - Register-sized objects vs. “large objects”
  - Cross-module optimization

# CONCLUSIONS

---

- System F  $\rightarrow$  TAL
  - We *can* have security and expressivity
  - Utilizes many PL techniques
    - Type-directed Compilation
  - Formalism omits many optimizations (other work)
- Future Work & Impact
  - Cyclone (low level, typed language)
    - (and then Rust)

**THANK YOU!**

# POLYMORPHIC CC – TWICE EXAMPLE

---

$\lambda^F$  source:

$twice = \Lambda\alpha. \lambda f:\alpha \rightarrow \alpha. \lambda x:\alpha. f(fx)$

$\lambda^K$  source:

$twice =$

$\lambda[\alpha](f:\tau_f, c:(\tau_f) \rightarrow void).$

  let  $twicef =$

$\lambda(x:\alpha, c':(\alpha) \rightarrow void).$

      let  $oncef = \lambda(z:\alpha). f(z, c')$  in

$f(x, oncef)$

  in

$c[] (twicef)$

where  $\tau_f = (\alpha, (\alpha) \rightarrow void) \rightarrow void$

# POLYMORPHIC CC

.....

$\lambda^F$  source:

$twice = \Lambda\alpha. \lambda f:\alpha \rightarrow \alpha. \lambda x:\alpha. f(fx)$

$\lambda^K$  source:

$twice =$

$\lambda[\alpha](f:\tau_f, c:(\tau_f) \rightarrow void).$

let  $twicef =$

$\lambda(x:\alpha, c':(\alpha) \rightarrow void).$

let  $oncef = \lambda(z:\alpha). f(z, c')$  in

$f(x, oncef)$

in

$c[](twicef)$

where  $\tau_f = (\alpha, (\alpha) \rightarrow void) \rightarrow void$

$\lambda^H$  translation:

letrec  $twice_{code}[\alpha](env:\langle \rangle, f:\tau_f, c:\exists\rho_3.\langle(\rho_3, \tau_f) \rightarrow void, \rho_3\rangle).$

let  $twicef = \text{pack } [\langle\tau_f\rangle, \langle twicef_{code}[\alpha], \langle f \rangle \rangle]$  as  $\tau_f$  in

let  $[\rho_3, c_{unpack}] = \text{unpack } c$  in

let  $c_{code} = \pi_1(c_{unpack})$  in

let  $c_{env} = \pi_2(c_{unpack})$  in

$c_{code}(c_{env}, twicef)$

$twicef_{code}[\alpha](env:\langle\tau_f\rangle, x:\alpha, c':\tau_{\alpha c}).$

let  $f = \pi_1(env)$  in

let  $oncef = \text{pack } [\langle\tau_f, \tau_{\alpha c}\rangle, \langle oncef_{code}[\alpha], \langle f, c' \rangle \rangle]$  as  $\tau_{\alpha c}$

let  $[\rho_1, f_{unpack}] = \text{unpack } f$  in

let  $f_{code} = \pi_1(f_{unpack})$  in

let  $f_{env} = \pi_2(f_{unpack})$  in

$f_{code}(f_{env}, x, oncef)$

$oncef_{code}[\alpha](env : \langle\tau_f, \tau_{\alpha c}\rangle, z : \alpha).$

let  $f = \pi_1(env)$  in

let  $c' = \pi_2(env)$  in

let  $[\rho_1, f_{unpack}] = \text{unpack } f$  in

let  $f_{code} = \pi_1(f_{unpack})$  in

let  $f_{env} = \pi_2(f_{unpack})$  in

$f_{code}(f_{env}, z, c')$

in ...

where  $\tau_f = \exists\rho_1.\langle(\rho_1, \alpha, \tau_{\alpha c}) \rightarrow void, \rho_1\rangle$

$\tau_{\alpha c} = \exists\rho_2.\langle(\rho_2, \alpha) \rightarrow void, \rho_2\rangle$