

Static Analysis of JavaScript Programs

Advisor ↪ **Dr Manas Thakur**
Presenter ↪ **Meetesh Kalpesh Mehta (23d0361)**, PhD2



Goal

construct an

Intermediate Representation for JavaScript

Goal

construct ~~an~~ a
general-purpose

Intermediate Representation for JavaScript

Goal

construct ~~an~~ a
general-purpose
analyzable

Intermediate Representation for JavaScript

Goal

construct an a
general-purpose
analyzable
executable

Intermediate Representation for JavaScript

Use Case



Amazon LLRT (Low Latency Runtime)

- **Experimental**
- **Lightweight**
- Address the growing demand for **fast and efficient serverless applications.**

Use Case

Amazon LLRT (Low Latency Runtime)

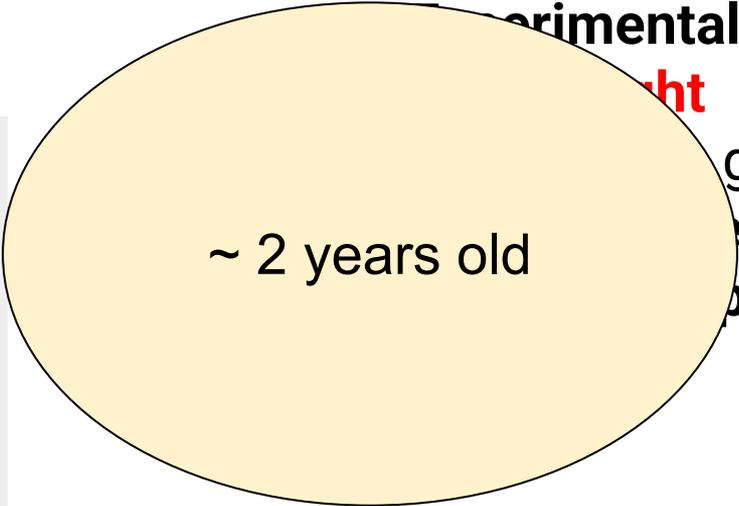
Experimental

Light

growing demand

efficient

applications.



~ 2 years old

Use Case



Amazon LLRT (Low Latency Runtime)

- **Experimental**
- **Lightweight**
- Address the growing demand for **fast and efficient serverless applications.**

Use Case

Amazon LLRT (Low Latency Runtime)

Experimental

quickjs-ng

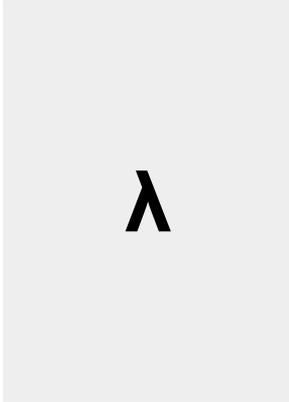
100-300KB Ram

no JIT

growing demand
efficient
applications.

λ

Use Case



Amazon LLRT (Low Latency Runtime)

- **Experimental**
- **Lightweight**
- Address the growing demand for **fast and efficient serverless applications.**

Use Case

Amazon LLRT (Low Latency Runtime)

Experimental

Light

growing demand

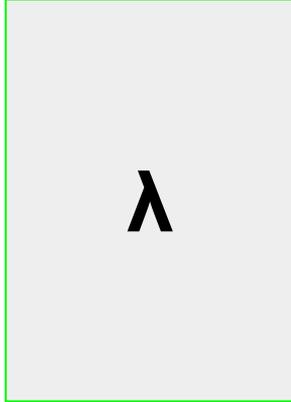
efficient

applications.

λ

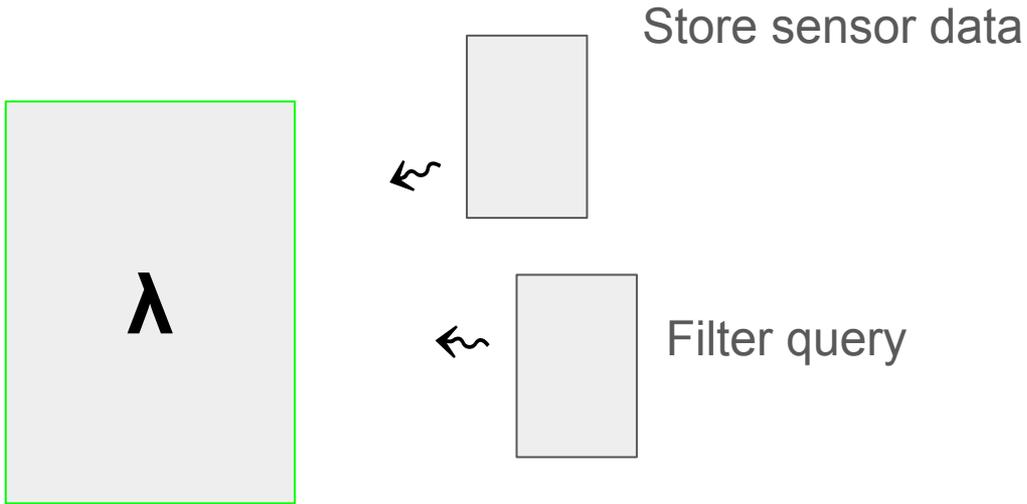
Why is this interesting?

Use Case

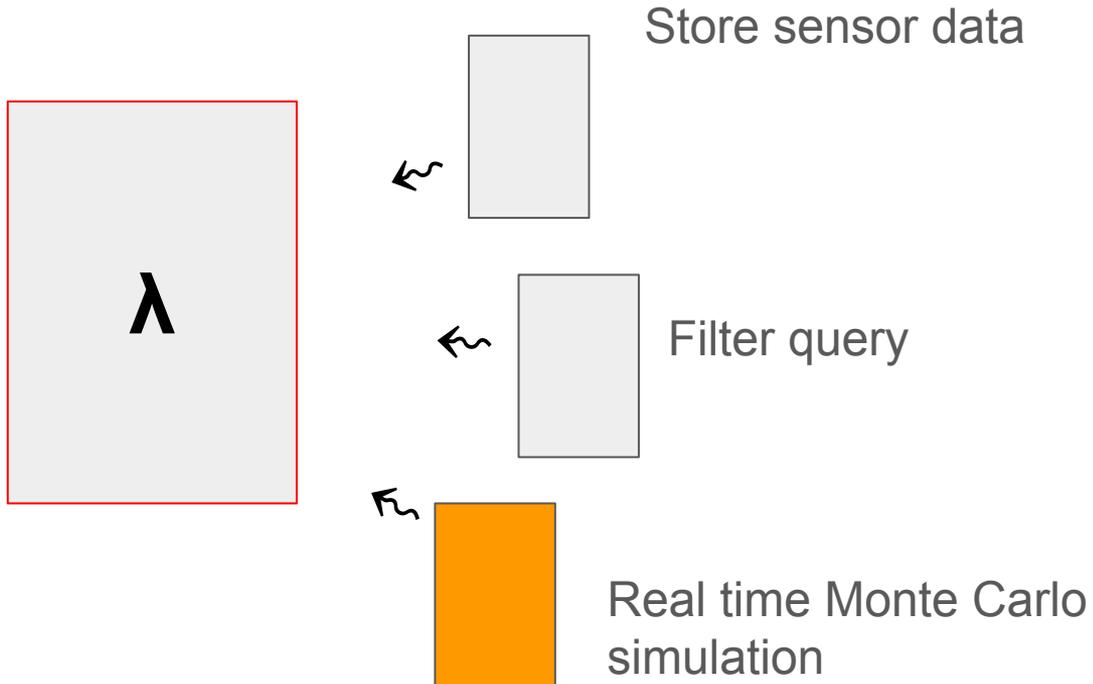


Store sensor data

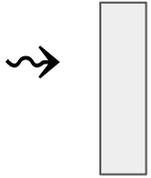
Use Case



Use Case



Hypothesising the limitation



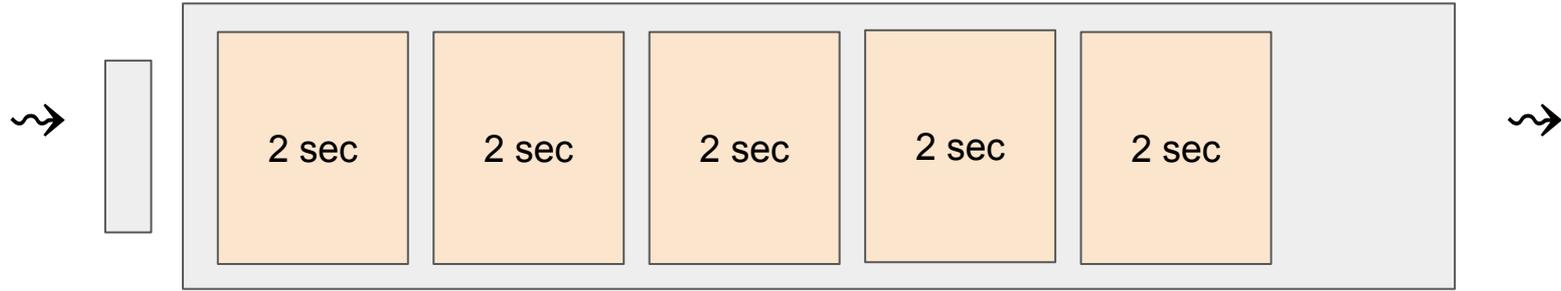
Couple hundred ms

Hypothesising the limitation



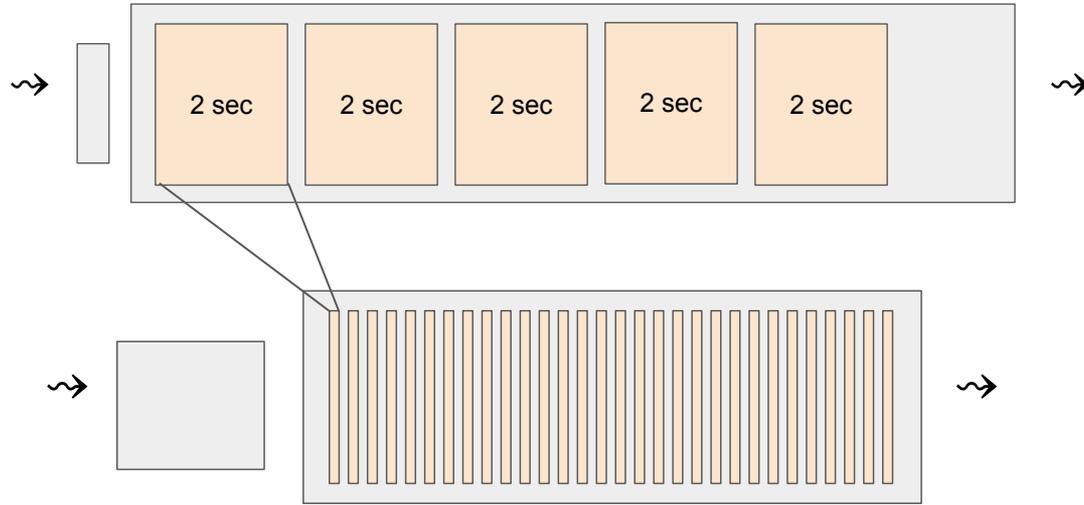
Couple hundred ms

Hypothesising the limitation



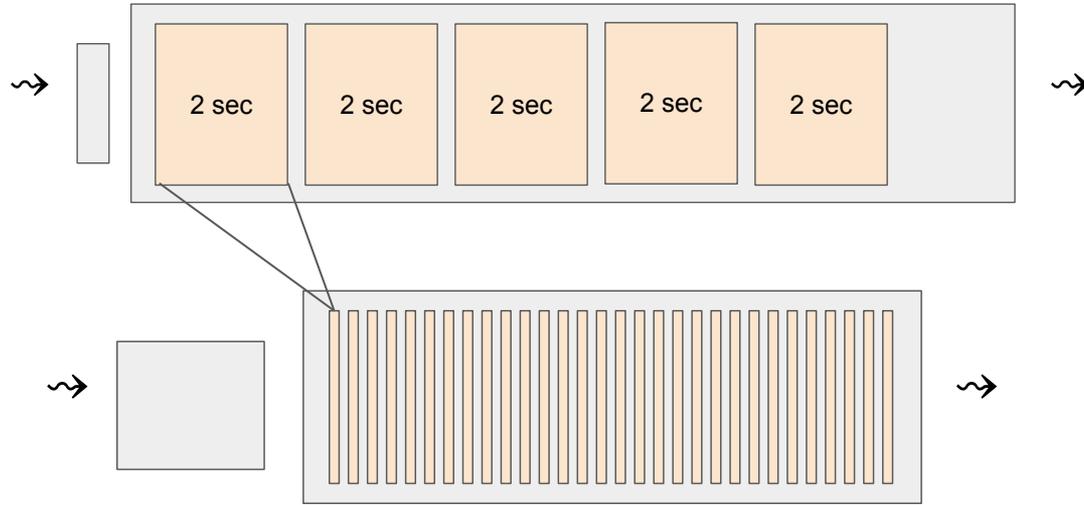
Couple hundred ms

Doing things Just-In-Time



Amortization of cost

The power of JIT



Amortization of cost

**Just In Time
Compilation**

Static Analysis
+
Optimized Code
Generation

Why no JIT?

Poor security (code region memory is sacred)

- Chrome on your phone says so!
- Apple won't allow it!

Why no JIT?

Poor security (code region memory is sacred)

Unpredictable (2s or 20s)

- **Amortization is not always guaranteed**
 - **Regression cases!!**
 - Most JITs have their WIP lists!

Why no JIT?

Poor security (code region memory is sacred)

Unpredictable (2s or 20s)

Not suitable for all workloads

- **High Throughput** ↑ >>> **Optimal Code** ↓

Our Pipeline

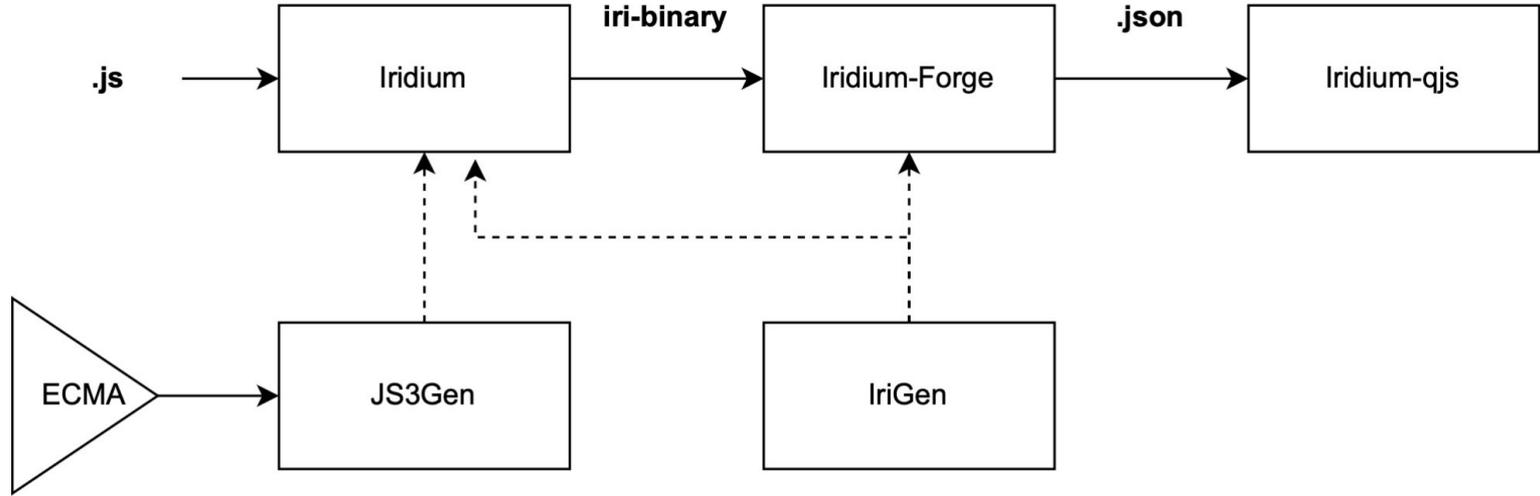


Figure 3.1: Overview of the Iridium architecture.

Our Pipeline

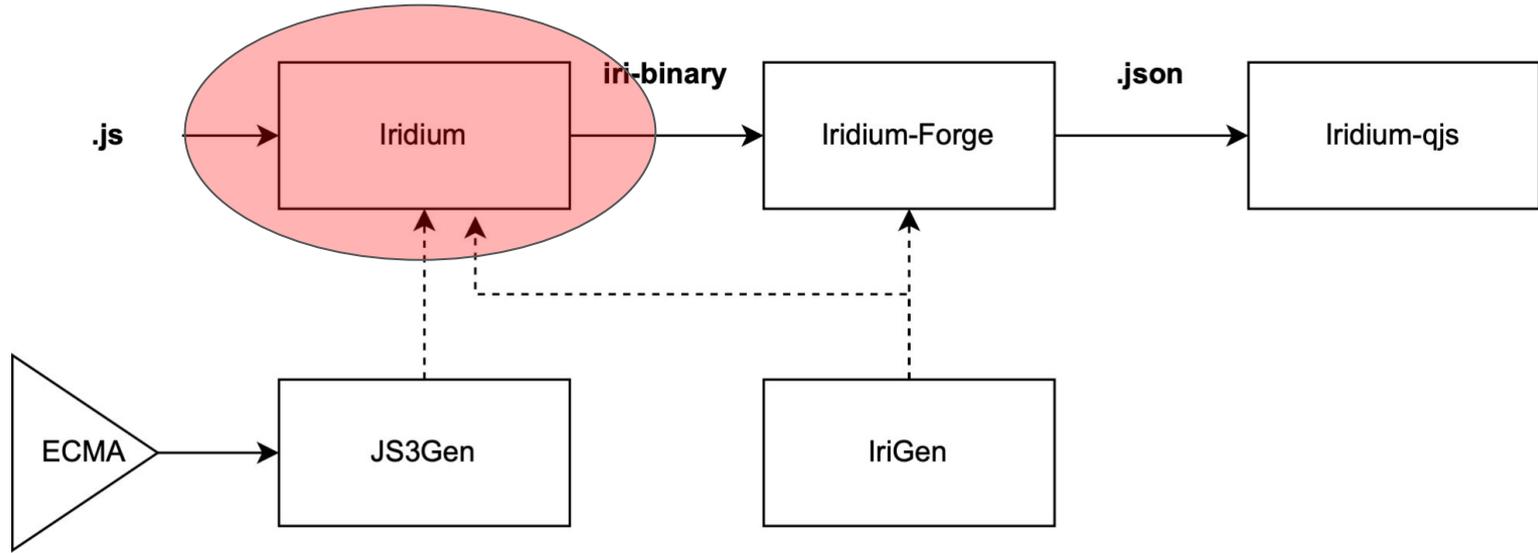


Figure 3.1: Overview of the Iridium architecture.

Our Pipeline

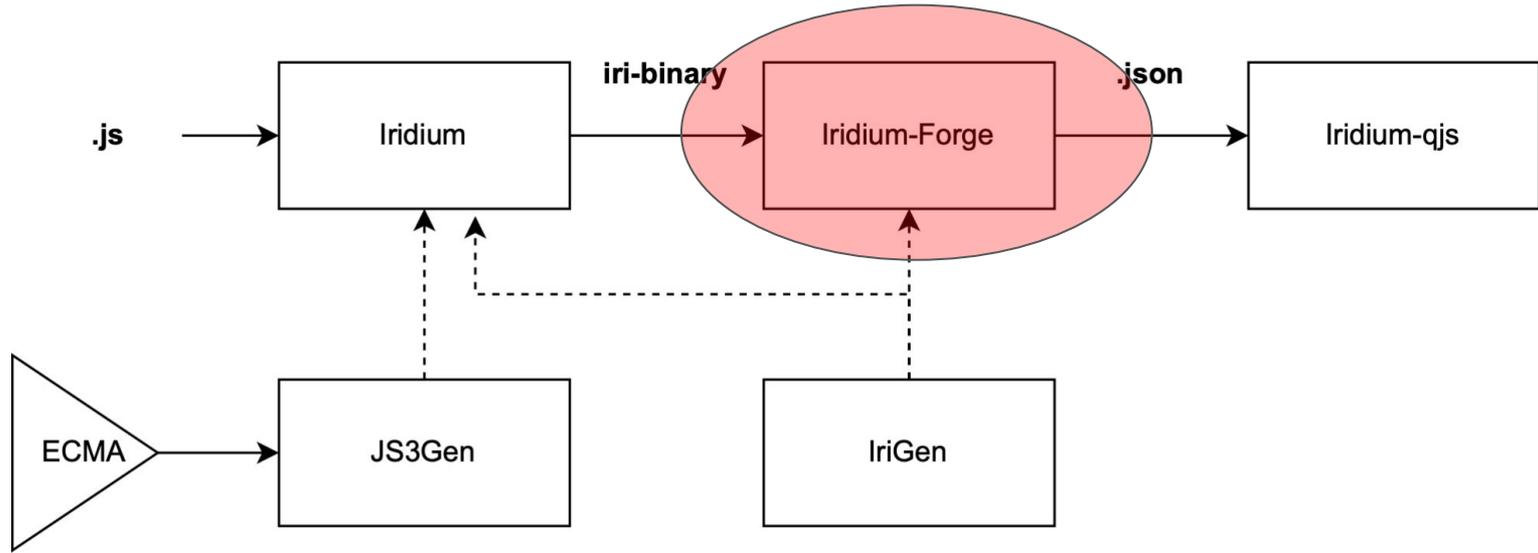


Figure 3.1: Overview of the Iridium architecture.

Our Pipeline

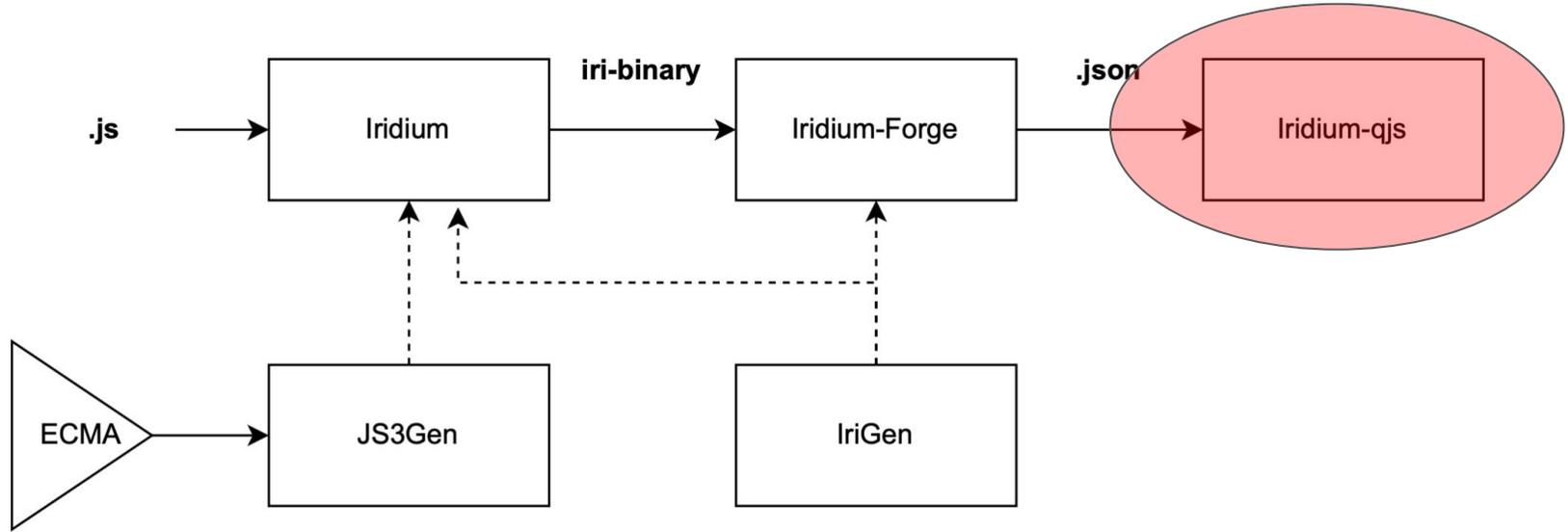


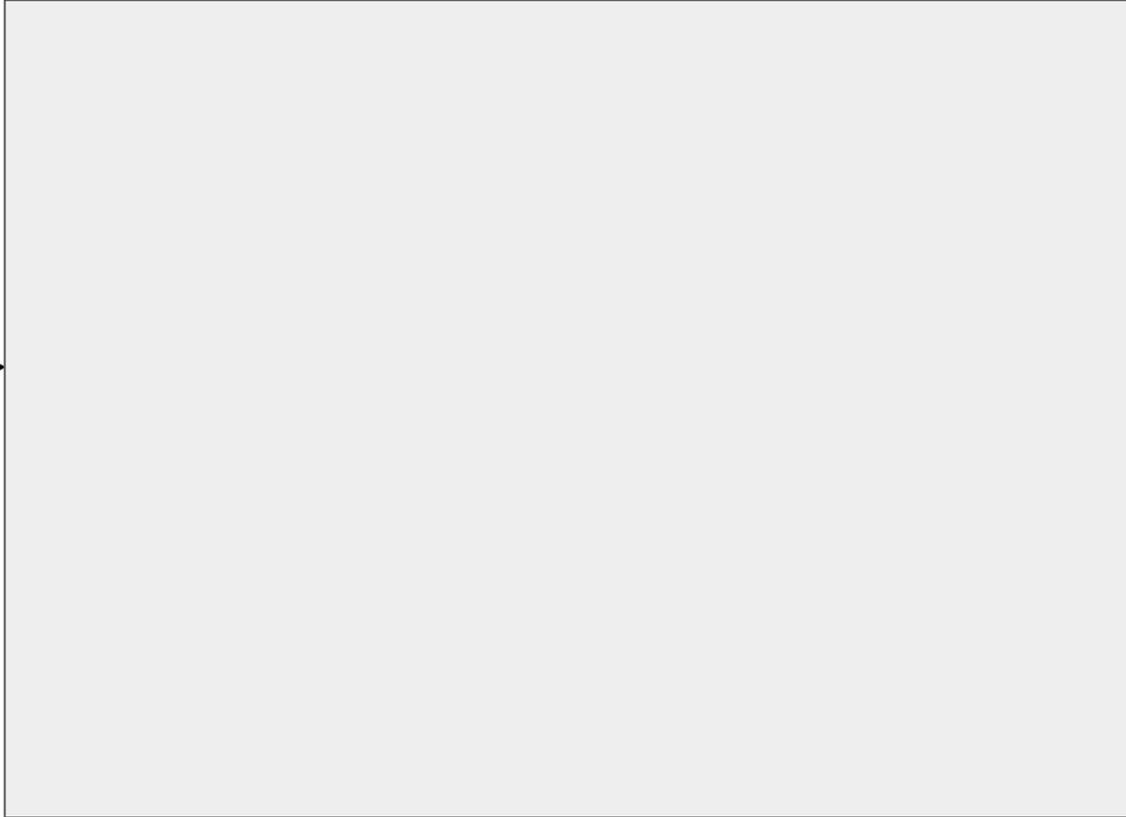
Figure 3.1: Overview of the Iridium architecture.

Frontend.

Just a loop right?

JS Source

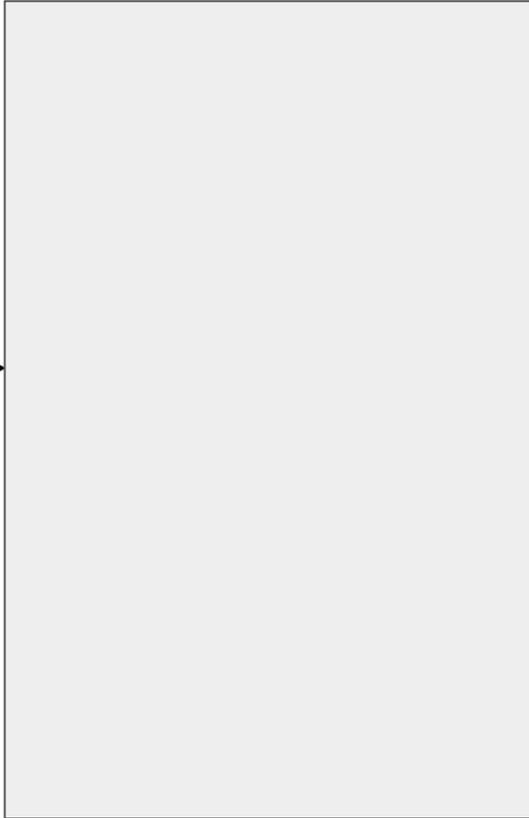
```
for (let a of [1,2,3]) {  
  let x = a;  
}
```



Just a loop right? **Wrong**

JS Source

```
for (let a of [1,2,3]) {  
  let x = a;  
}
```



Execution Virtual Machine

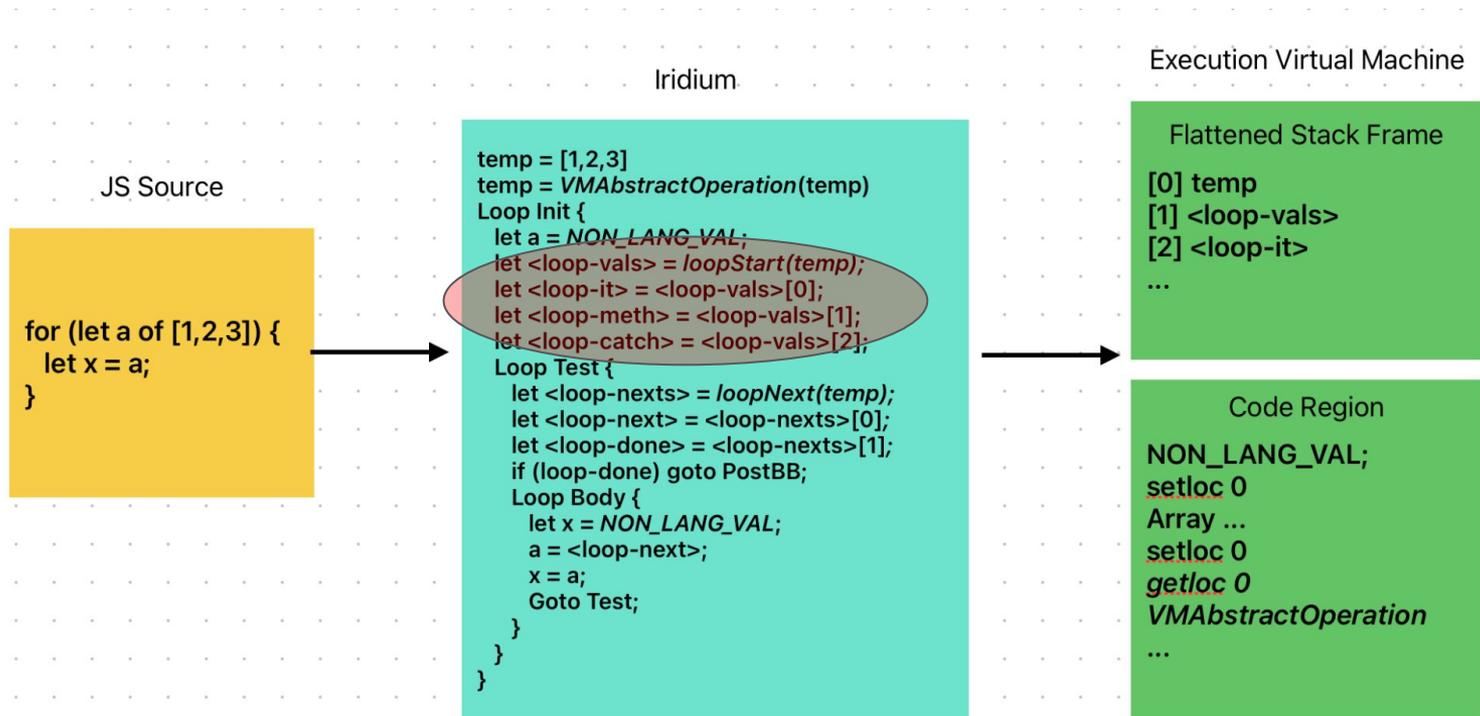
Flattened Stack Frame

```
[0] temp  
[1] <loop-vals>  
[2] <loop-it>  
...
```

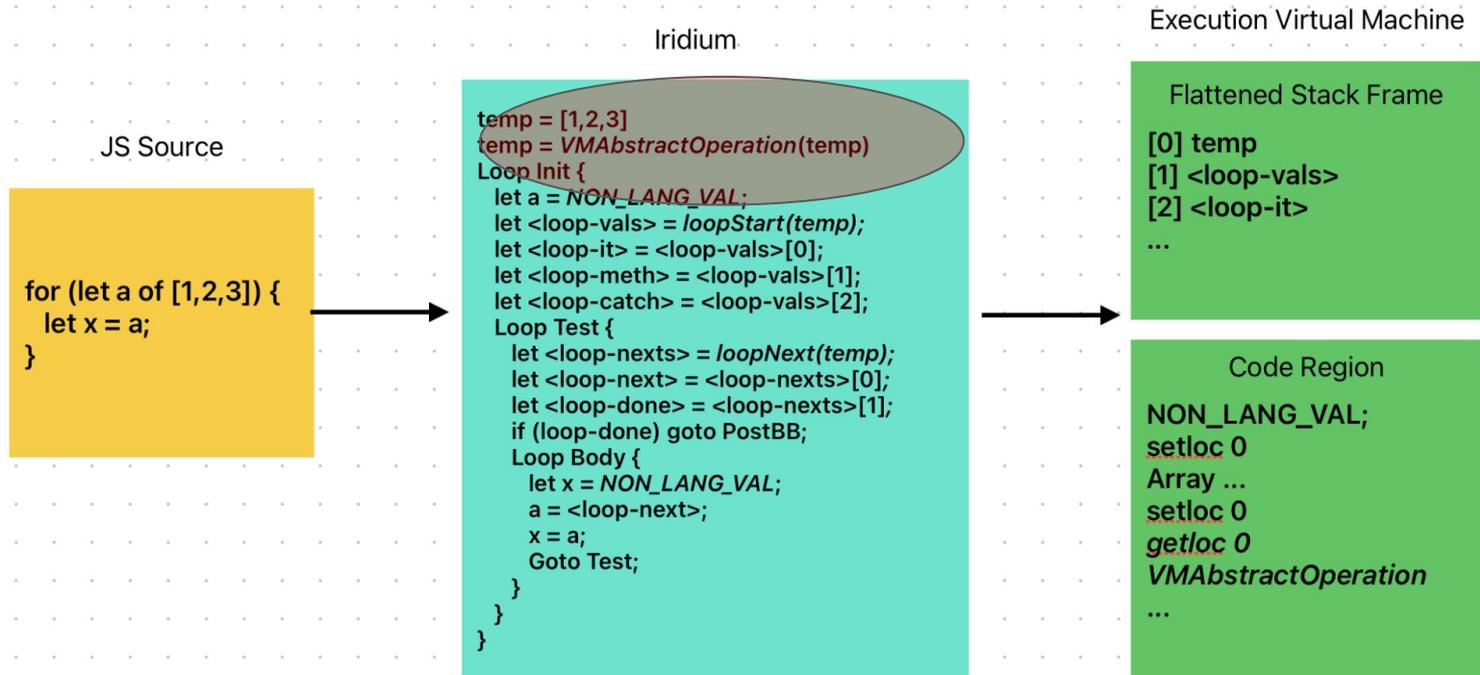
Code Region

```
NON_LANG_VAL;  
setloc 0  
Array ...  
setloc 0  
getloc 0  
VMAbstractOperation  
...
```

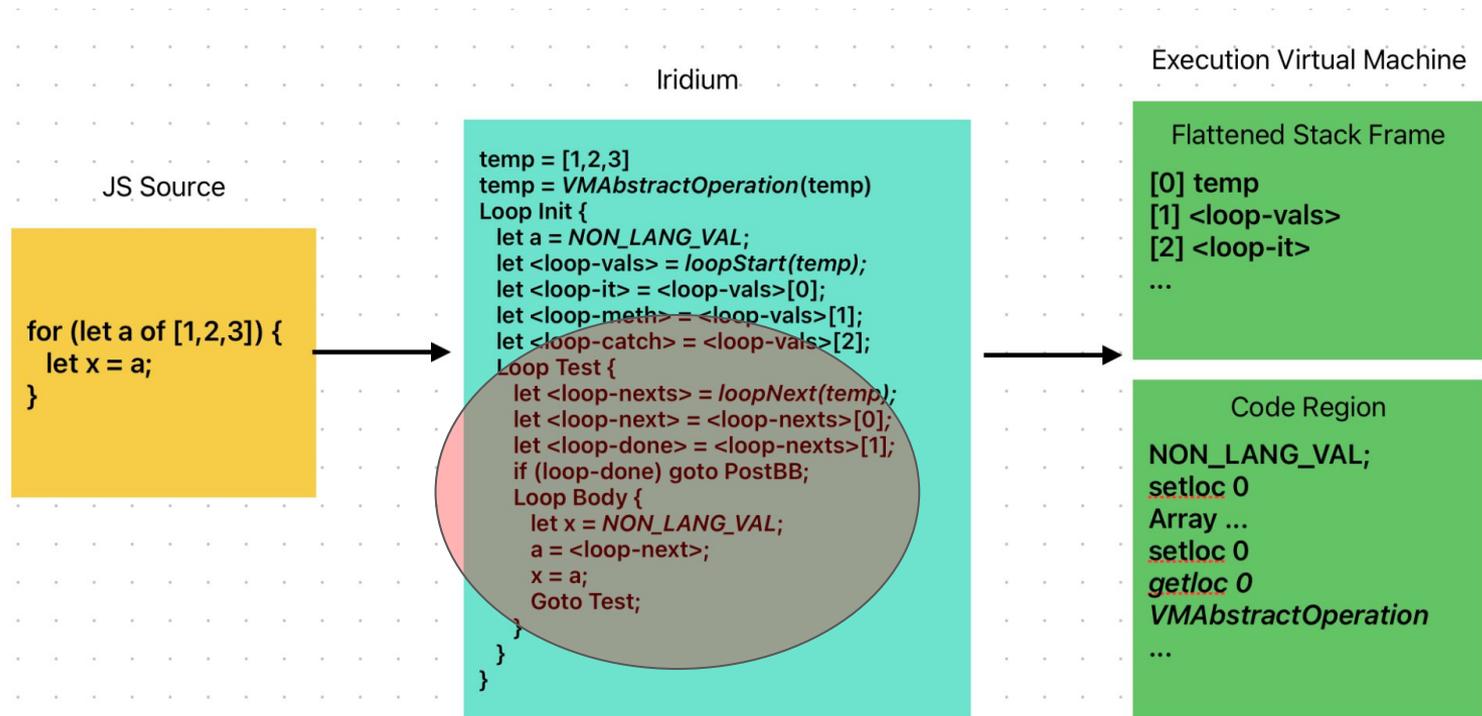
Try Catch Block, Implicit Abstract Operations and new evaluation scopes



Try Catch Block, Implicit Abstract Operations and new evaluation scopes



Try Catch Block, Implicit Abstract Operations and **new** evaluation scopes



Iridium SEXP Format

IridiumSEXP

String : IridiumPrimitive

Args

A generic S-Expression has this format.

Tag : Denotes the SEXP kind

Flags: Quirks on the SEXP

Args: The children of the node

Iridium Primitives

number | boolean | string | null;

`.js` \rightsquigarrow `.js3` \rightsquigarrow `.iri(non-exe)` \rightsquigarrow `.iri(exe)` \rightsquigarrow `.json` \rightsquigarrow bytecode

```
let a = 12;  
{  
  console.log(a);  
  let a = 13;  
}
```

What's the output?

`.js` \rightsquigarrow `.js3` \rightsquigarrow `.iri(non-exe)` \rightsquigarrow `.iri(exe)` \rightsquigarrow `.json` \rightsquigarrow bytecode

```
let a = 12;  
{  
  console.log(a);  
  let a = 13;  
}
```

```
let a = 12;  
{  
  let js3$1 = console.log(a);  
  let a = 13;  
}
```

```
./iridium js3 --tout tests/basic/ex1.mjs
```

Running Example

.js ~> .js3 ~> .iri(non-exe) ~> .iri(exe) ~> .json ~> bytecode

```
let a = 12;
{
  let js3$1 = console.log(a);
  let a = 13;
}
```

```
TMo0N nqqqBBH]
AJ7nM8/□vqqqqqq@}qqq!~
(Ma
  Ppff}qq}qqB"g餵q>IN#+HMqĀ_P

A_s5RZqHqj0q|1x(XqLZqf^.-i:AHqCqqGq_ArqqHNq
`q':]q4qκi

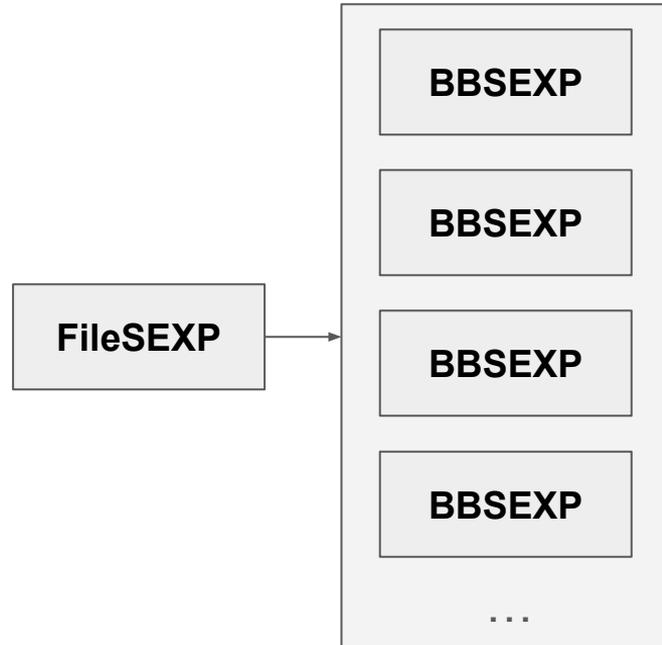
qqqqqqqqmYsqqNجس:
qr\gqNtqWqqqqqvqqmqQ&f1qaqq
8q:vqqqRoqvqWq}KgqC(q]q4/q&f*qx(q#q:2qqDq]
~D&qnqq%^qqq-GJ8srgqS(q(rq;
qqqqq|qq7]{tJ/qZqq8qqqqqqgqIq-`1qqq
q7qqq2qkqtq醜qqqqq<qPq qqbb^ PqqqCcq3q)q-Wqrquq
(qqqqqqqCyqqeqq]T`qqq/qqq1cqqq1qq_8?8gqq28
Nbv1Yqqq'q/q4qKq qqLqHq q`qS&qqqqiq>"q;
qqq

lyqqqqqqqq9Rq7@qxq;" U]qqdqR68[q~qqqT;0!q+qqqqqqqqeGD?
q)yoqQNqq견
  Q\L2qEq=uqq,5dqqqq□qDq* Cdq|qqk:
q8qVI?qU]<n}^qg?#|qq"cqbbqqq>qqq
```

```
./iridium iri --debugIri . ./tests/basic/test1.mjs
```

Running Example

`.js` \rightsquigarrow `.js3` \rightsquigarrow `.iri(non-exe)` \rightsquigarrow `.iri(exe)` \rightsquigarrow `.json` \rightsquigarrow bytecode



Running Example

.js ↷ .js3 ↷ **.iri(non-exe)** ↷ .iri(exe) ↷ .json ↷ bytecode

BB0:

```
{this, VSYSCALL(9)}  
goto (this) ? 1 : 2
```

BB1:

```
Return undefined
```

BB2:

```
{<module_meta>, VSYSCALL(6)}  
{a, 12}  
goto 3
```

BB3:

```
{ccallCallee$2, console.log}  
CallSite({console}, {ccallCallee$2}, {a})  
{a, 13}  
goto 4
```

BB4:

```
AsyncReturn undefined
```

.js ↗ .js3 ↗ .iri(non-exe) ↗ .iri(exe) ↗ .json ↗ bytecode

BB0:

```
{this, VSYSCALL(9)}  
goto (this) ? 1 : 2
```

BB1:

Return undefined

BB2:

```
{<module_meta>, VSYSCALL(6)}  
{a, 12}  
goto 3
```

BB3:

```
{ccallCallee$2, console.log}  
CallSite({console}, {ccallCallee$2}, {a})  
{a, 13}  
goto 4
```

BB4:

AsyncReturn undefined

.js ↗ .js3 ↗ .iri(non-exe) ↗ .iri(exe) ↗ .json ↗ bytecode

BB0:

```
{this, VSYSCALL(9)}  
goto (this) ? 1 : 2
```

BB1:

```
Return undefined
```

BB2:

```
{<module_meta>, VSYSCALL  
{a, 12}  
goto 3
```

BB3:

```
...lee$2, cor  
...sole},
```

**JSImplicitBinding
DeclarationSEXP**



.js ↷ .js3 ↷ .iri(non-exe) ↷ .iri(exe) ↷ .json ↷ bytecode

BB0:

```
{this, VSYSCALL(9)}  
goto (this) ? 1 : 2
```

BB1:

Return undefined

BB2:

```
{<module_meta>, VSYSCALL(6)}  
{a, 12}  
goto 3
```

BB3:

```
{ccallCallee$2, console.log}  
CallSite({console}, {ccallCallee$2}, {a})  
{a, 13}  
goto 4
```

BB4:

AsyncReturn undefined

.js ↷ .js3 ↷ .iri(non-exe) ↷ .iri(exe) ↷ .json ↷ bytecode

BB0:

```
{this, VSYSCALL(9)}  
goto (this) ? 1 : 2
```

BB1:

```
Return undefined
```

BB2:

```
{<module_meta>, VSYSCALL  
{a, 12}  
goto 3
```

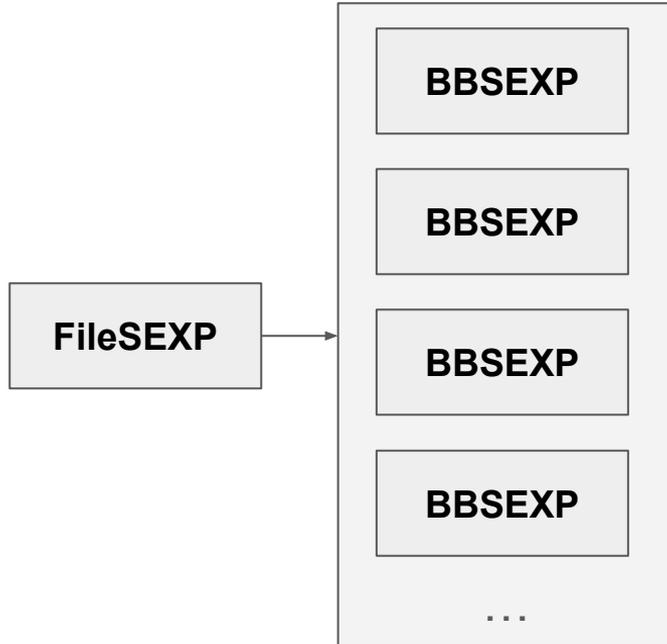
BB3:

```
!lee$2, console.log  
console},
```

**JSExplicitBinding
DeclarationSEXP**



.js ~> .js3 ~> .iri(non-exe) ~> .iri(exe) ~> .json ~> bytecode

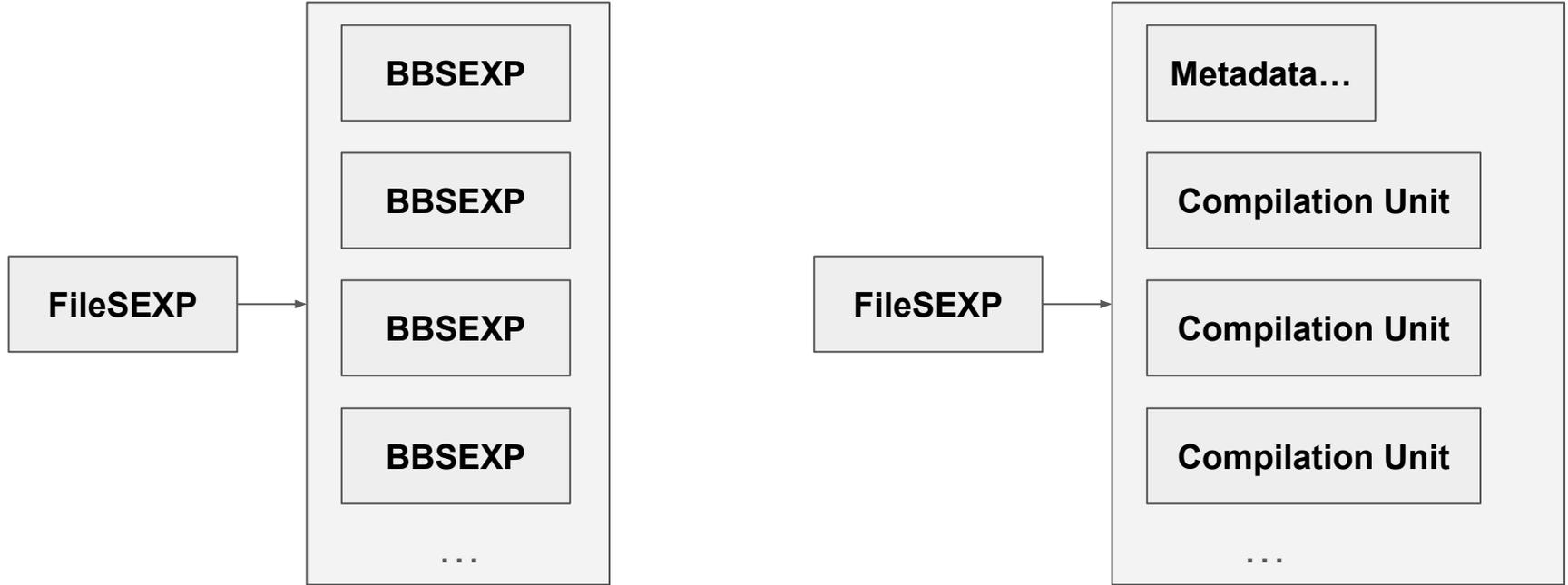


- Explicit control flow
- Transitional Nodes

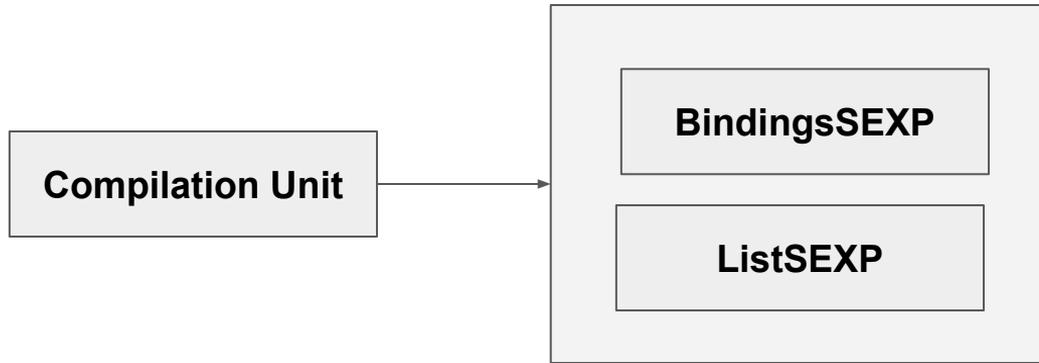
- No logical stack frames
- No closure grouping
- Incomplete semantics
 - Decorators
 - Stack Balancing

Forge.

Structural Separation happens in this phase



Structural Separation happens in this phase



BindingsSEXP is the logical stack frame

ListSEXP is a homogenous list of BBs

Structural Separation happens in this phase

```
let bar = () => {  
  console.log(foo);  
}  
let foo = 12;  
bar()
```

```
foo = NUBD  
bar = NUBD  
bar = Lambda@0  
foo = 12;  
bar()
```

```
Locals:  
  [0]bar  
  [1]foo  
Remote:
```

```
console.log(foo)
```

```
Locals:  
Remote:  
  [0][1]foo
```

Structural Separation happens in this phase

```
let bar = () => {  
  console.log(foo);  
}  
let foo = 12;  
bar()
```

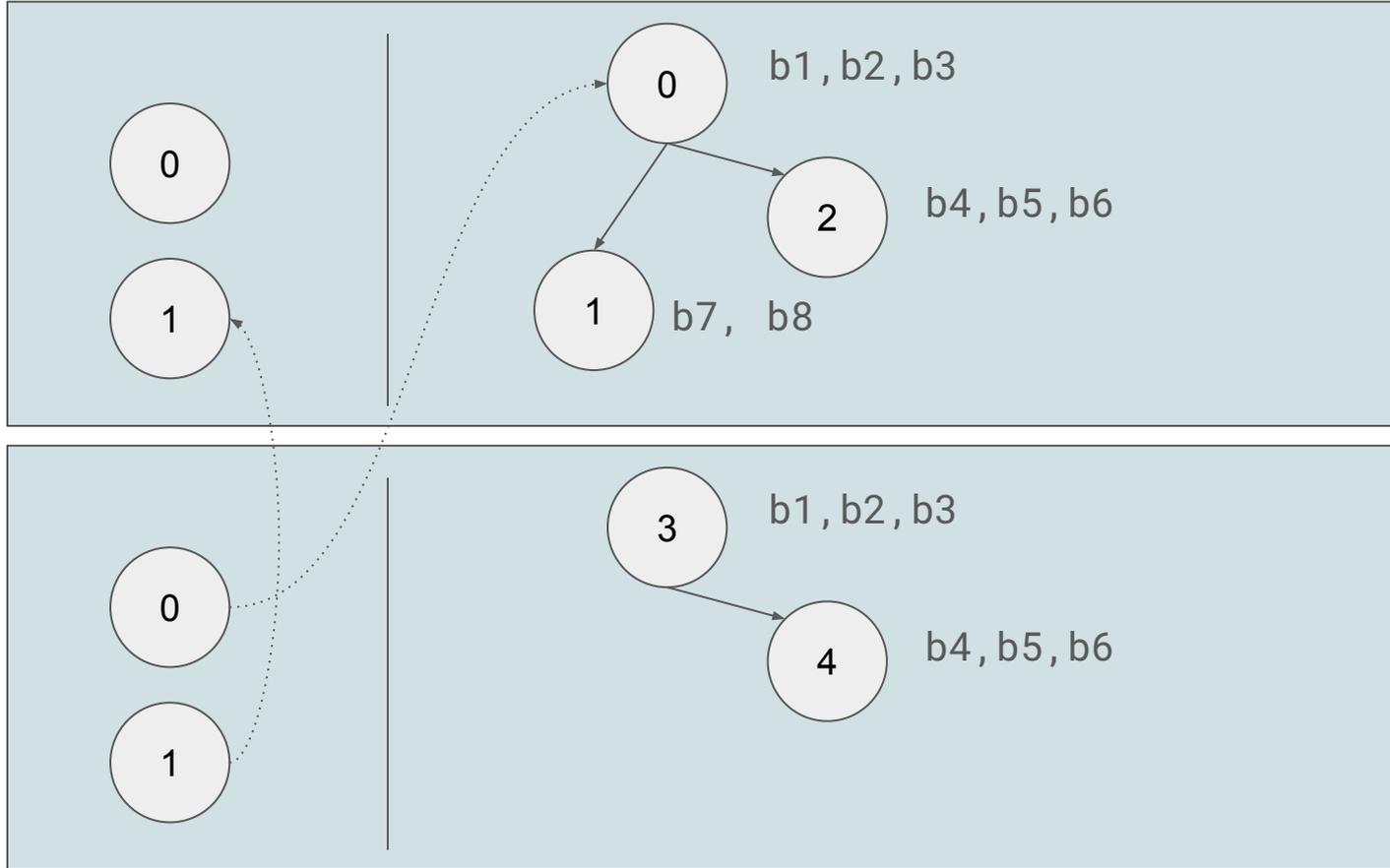
```
foo = NUBD  
bar = NUBD  
bar = Lambda@0  
foo = 12;  
bar()
```

```
Locals:  
  [0]bar  
  [1]foo  
Remote:
```

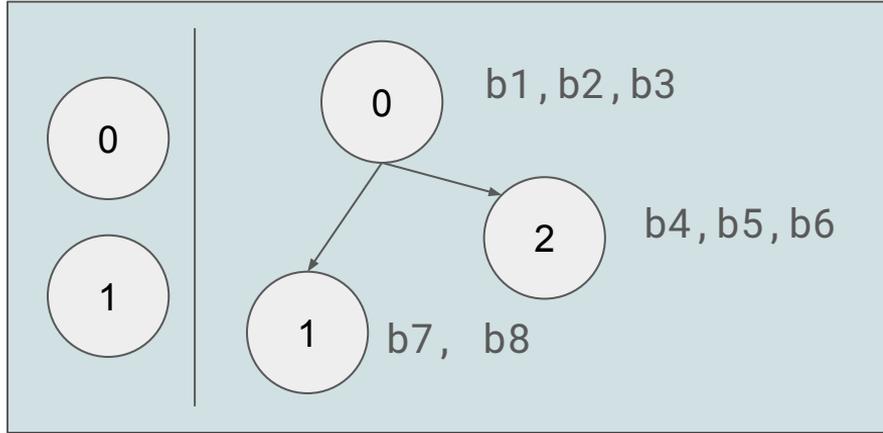
```
console.log(foo)
```

```
Locals:  
Remote:  
  [0][1]foo
```

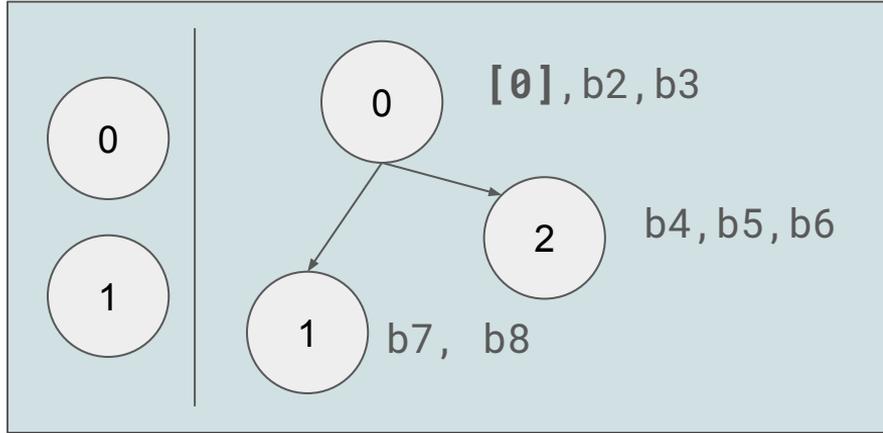
Logical Stack Frames



Logical Stack Frames - Flattening

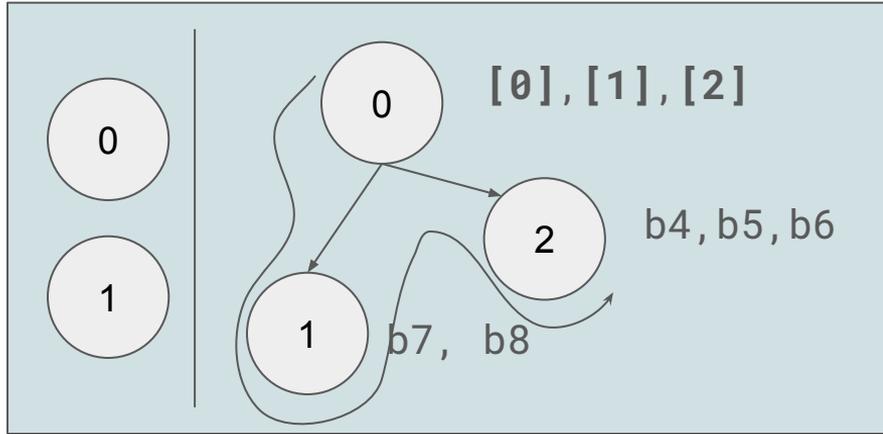


Logical Stack Frames - Flattening



$[0]$ b_1 $[-1]$

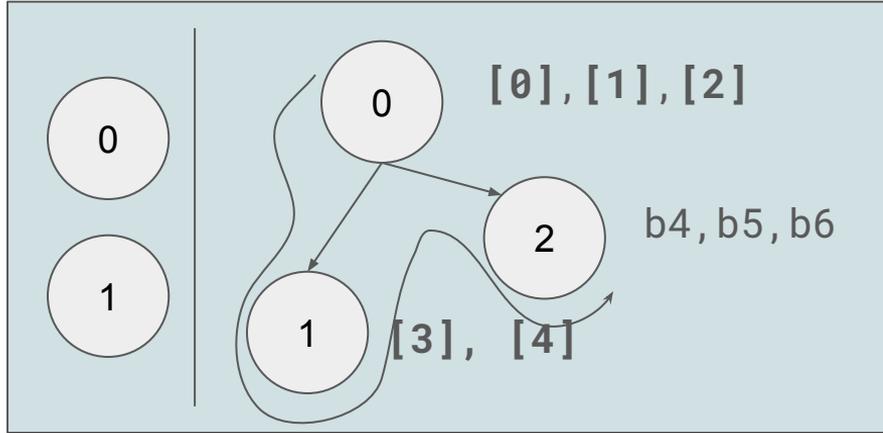
Logical Stack Frames - Flattening



[0]	b1	[-1]
[1]	b2	[0]
[2]	b3	[1]

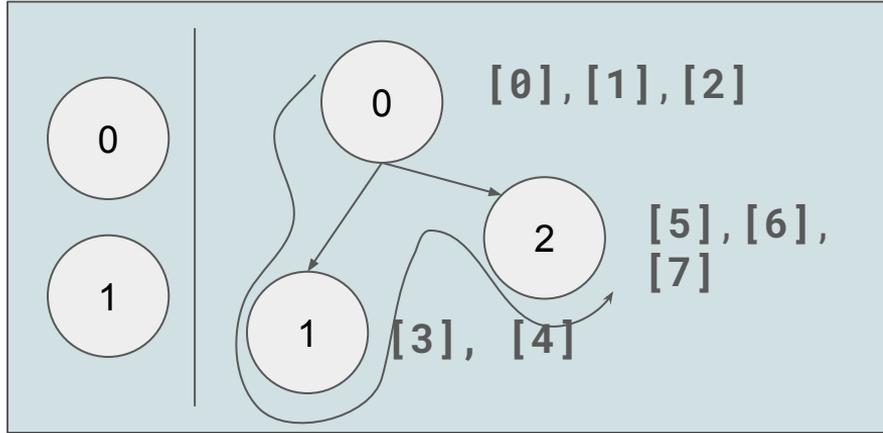
Preorder traversal of the scope tree

Logical Stack Frames - Flattening



[0]	b1	[-1]
[1]	b2	[0]
[2]	b3	[1]
[3]	b7	[2]
[4]	b8	[3]

Logical Stack Frames - Flattening

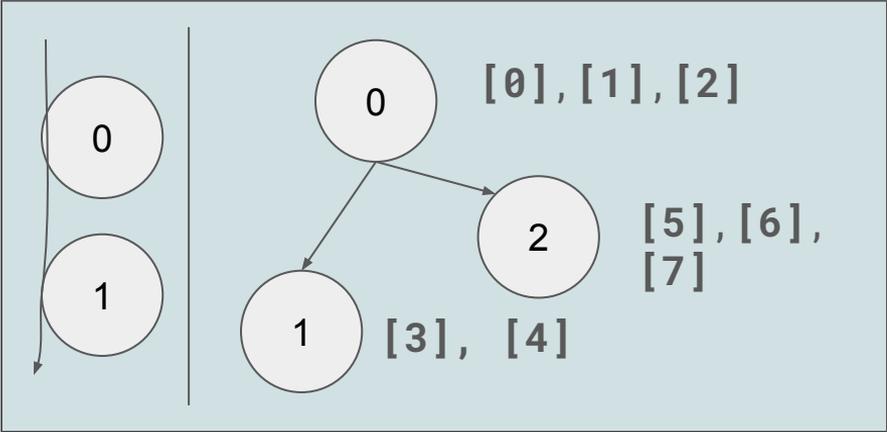


[0] b1 [-1]
[1] b2 [0]
[2] b3 [1]

[3] b7 [2]
[4] b8 [3]

[5] b4 [2]
[6] b5 [5]
[7] b6 [6]

Logical Stack Frames - Flattening



[0] b1 [-1]
[1] b2 [0]
[2] b3 [1]

[3] b7 [2]
[4] b8 [3]

[5] b4 [2]
[6] b5 [5]
[7] b6 [6]

[0] loc|remote [IDX]
[1] loc|remote [IDX]

`.js` \rightsquigarrow `.js3` \rightsquigarrow `.iri(non-exe)` \rightsquigarrow `.iri(exe)` \rightsquigarrow `.json` \rightsquigarrow bytecode

```
let a = 12;  
{  
  console.log(a);  
  let a = 13;  
}
```

What's the output?

.js ↷ .js3 ↷ .iri(non-exe) ↷ **.iri(exe)** ↷ .json ↷ bytecode

BB0:

```
[0]-[0] = NBUD
[0] = VSYSCALL(9) // this
If (this) goto 1
[1] = VSYSCALL(9) //<module_meta>
[0]-[0] = 12
[2] = NUBD
[3] = NUBD
[4] = NUBD
[2] = console.log
[3] = call (console, [2], [4])
[4] = 13
AsyncReturn undefined
```

BB1:

Return undefined

.js ↷ .js3 ↷ .iri(non-exe) ↷ **.iri(exe)** ↷ .json ↷ bytecode

BB0:

```
a = NBUD
this = VSYSCALL(9)
If (this) goto 1
<module_meta> = VSYSCALL(6)
a = 12
ccallCallee$2 = NUBD
js3$1 = NUBD
a = NUBD
ccallCallee$2 = console.log
js3$1 = call(console, ccallCallee$2, a)
a = 13
AsyncReturn undefined
```

BB1:

Return undefined

.js ↷ .js3 ↷ .iri(non-exe) ↷ **.iri(exe)** ↷ .json ↷ bytecode

BB0:

```
[0]-[0] = NBUD // remote a
[0] = VSYSCALL(9) // this
If (this) goto 1
[1] = VSYSCALL(9) //<module_meta>
[0]-[0] = 12 // remote a = 12
[2] = NUBD
[3] = NUBD
[4] = NUBD // local a
[2] = console.log
[3] = call (console, [2], [4])
[4] = 13 // local a = 12
AsyncReturn undefined
```

BB1:

Return undefined

```
[0] this [-1]
[1] <module_meta> [0]
[2] ccallCallee$2 [1]
[3] js3$1 [2]
[4] a [3]
```

```
[0] a:loc [0]
```

Running Example

.js ↷ .js3 ↷ .iri(non-exe) ↷ **.iri(exe)** ↷ .json ↷ bytecode

BB0:

```
[0]-[0] = NBUD // remote a
[0]     = VSYSCALL(9) // this
If (this) goto 1
[1]     = VSYSCALL(9) //<module_meta>
[0]-[0] = 12 // remote a = 12
[2]     = NUBD
[3]     = NUBD
[4]     = NUBD // local a
[2]     = console.log
[3]     = call (console, [2], [4])
[4]     = 13 // local a = 12
AsyncReturn undefined
```

BB1:

Return undefined

```
[0] this [-1]
[1] <module_meta> [0]
[2] ccallCallee$2 [1]
[3] js3$1 [2]
[4] a [3]
```

```
[0] a:loc [0]
```

Running Example

.js ↪ .js3 ↪ .iri(non-exe) ↪ **.iri(exe)** ↪ .json ↪ bytecode

NUBD = No Use Before Def
This will cause a runtime error

BB0:

[0]-[0]

[0]

If (this)

[1]

[0]-[0]

[2]

= NUBD

[3]

= NUBD

[4]

= NUBD // local a

[2]

= console.log

[3]

= call (console, [2], [4])

[4]

= 13 // local a = 12

AsyncReturn undefined

[0] this [-1]

[1] <module_meta> [0]

[2] ccallCallee\$2 [1]

[3] js3\$1 [2]

[4] a [3]

[0] a:loc [0]

Backend.

.js ↷ .js3 ↷ .iri(non-exe) ↷ .iri(exe) ↷ **.json** ↷ bytecode

```
{
  "version": "0.9a",
  "absoluteFilePath": "/home/meetesh/wd/Iridium/tests/basic/ex1.mjs",
  "iridium": [
    "File",
    [
      [
        "List",
        [],
        [
          [
            "TYPE",
            "ModuleRequest"
          ]
        ]
      ]
    ],
    ...
  ]
}
```

Running Example

Parsing

- Load and parse the Iridium code
- **cJSON library** is used to load the Iridium code into memory
- We haven't set a binary format (yet).
- Empirical results say that this part is not very costly (yet), so its low priority currently.

Instruction Selection

- We have a very basic instruction selection currently.
 - **Lot of room for improvement.**
- Independent problem from Iridium analysis/optimization.
 - We are considering introducing new intrinsics for **special cases.**

.js ↪ .js3 ↪ .iri(non-exe) ↪ .iri(exe) ↪ .json ↪ **bytecode**

```
ex1.mjs:1:1: function: <null>
  mode: strict
  locals:
    0: const this
    1: const <module_meta>
    2: var ccallCallee$2 [level:2 next:1]
    3: var js3$1 [level:2 next:2]
    4: var a [level:2 next:3]
  closure vars:
    0: a local:loc0 var
  stack_size: 3
  byte_code_len: 93
  opcodes: 30
```

.js ↪ .js3 ↪ .iri(non-exe) ↪ .iri(exe) ↪ .json ↪ **bytecode**

```
push_const      0      ; uninitialized
put_var_ref     0      ; a
push_this
put_loc         0      ; this
get_loc_check   0      ; this
if_true        91
special_object  6
put_loc         1      ; "<module_meta>"
push_i8        12
put_var_ref     0      ; a
push_const     1      ; uninitialized
put_loc         2      ; ccallCallee$2
push_const     2      ; uninitialized
put_loc         3      ; js3$1
```

...

.js ↪ .js3 ↪ .iri(non-exe) ↪ .iri(exe) ↪ .json ↪ **bytecode**

```
push_const    0    ; uninitialized  
put_var_ref   0    ; a
```

Possibly unhandled promise rejection: ReferenceError: a is not initialized
at <anonymous> (/home/meetesh/wd/Iridium/tests/basic/ex1.mjs:1:1)

```
put_loc      2    ; uninitialized  
push_const   2    ; uninitialized  
put_loc      3    ; js3$1
```

...

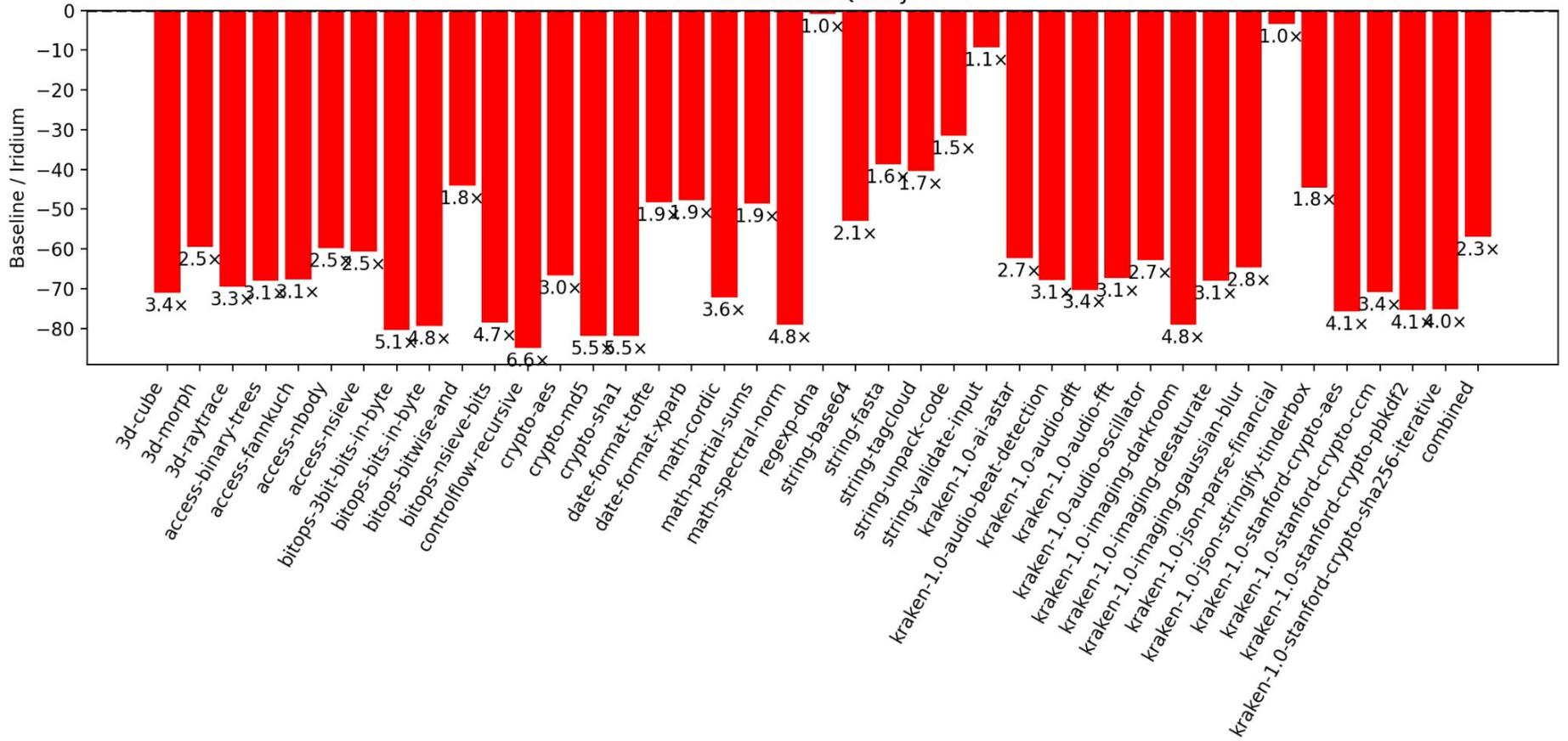
Conclusion.

Project Statistics

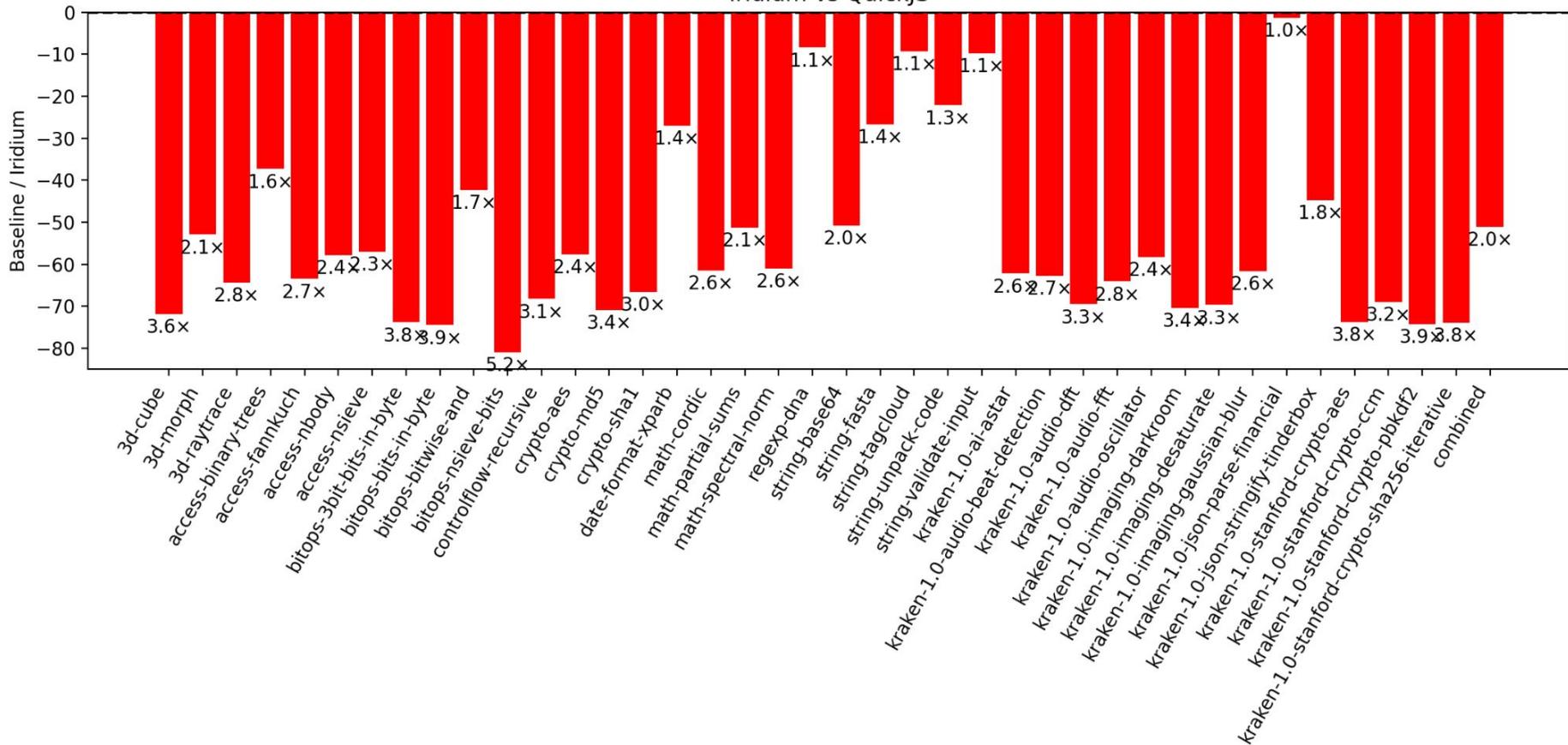
Project	LOC	Languages	Status
Iridium	~50K	TypeScript	Stable, WIP
JS3Gen	~1K	TypeScript	Mostly Stable
Iridium-Forge	~9K	C++	WIP
IriGen	~200	TypeScript	Mostly Stable
Iridium-qjs (patch)	~5K	C/C++	WIP, one stable release

Table 3.1: Breakdown of Iridium projects by size, language, and stability.

Iridium vs QuickJS



Iridium vs QuickJS



Immediate Priority

- Make Iridium code execute **reasonably faster** than baseline.
 - We believe **traditional optimization passes** will help us achieve this goal.
 - Traditional optimization passes require **augmentation** to support closure semantics and modelling of **read/write barriers**.

Future

- High level analysis/optimization
 - **3-4 months**
- Analysis of asynchronous primitives (security/parallelism)
 - **6-8 months**
- More frontends, Python? Lua? Scheme?
 - **6-8 months**
- Long term goals
 - **1-2 yrs ~ More Backends**