# Generating Optimized Code with GloballSel

Or: GloballSel going beyond "it works"



### Agenda

- What is GloballSel?
- GloballSel Combiner and Helpers
- Testing and Debugging
- Declarative Combiner



# But first...



### History

- In 2017, we got GloballSel fully working for our target
  - Fast compile time, but codegen quality was significantly lower



- In 2017, we got GloballSel fully working for our target
  - Fast compile time, but codegen quality was significantly lower
- Added several new features to improve codegen quality



- In 2017, we got GloballSel fully working for our target
  - Fast compile time, but codegen quality was significantly lower
- Added several new features to improve codegen quality
- By 2019, the codegen quality has improved



# Apple GPU Compiler Uses GloballSel

GloballSel is a new instruction selection framework

- GloballSel is a new instruction selection framework
- Supports more global optimization (e.g. match across BasicBlocks)

- GloballSel is a new instruction selection framework
- Supports more global optimization (e.g. match across BasicBlocks)
- More flexible
  - From the speed of FastISel to the quality of SelectionDAGISel

- GloballSel is a new instruction selection framework
- Supports more global optimization (e.g. match across BasicBlocks)
- More flexible
  - From the speed of FastISel to the quality of SelectionDAGISel
- Easier to understand, maintain, and test

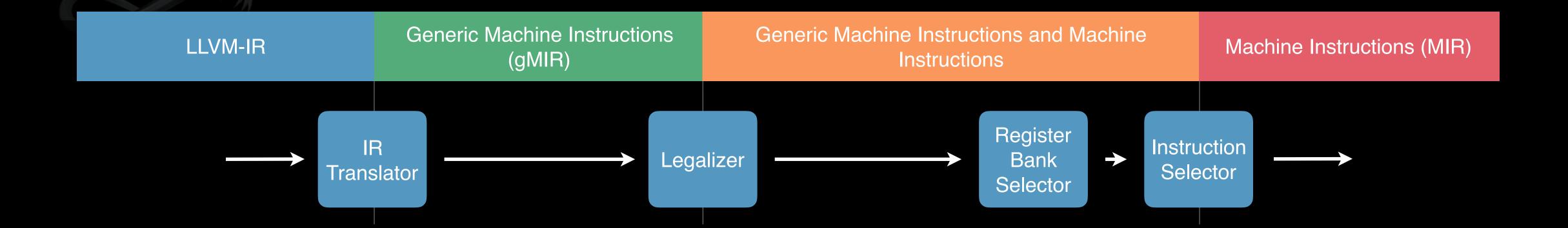
- GloballSel is a new instruction selection framework
- Supports more global optimization (e.g. match across BasicBlocks)
- More flexible
  - From the speed of FastISel to the quality of SelectionDAGISel
- Easier to understand, maintain, and test
- Keeps all state in the Machine IR (MIR)

- GloballSel is a new instruction selection framework
- Supports more global optimization (e.g. match across BasicBlocks)
- More flexible
  - From the speed of FastISel to the quality of SelectionDAGISel
- Easier to understand, maintain, and test
- Keeps all state in the Machine IR (MIR)

A Proposal for Global Instruction Selection

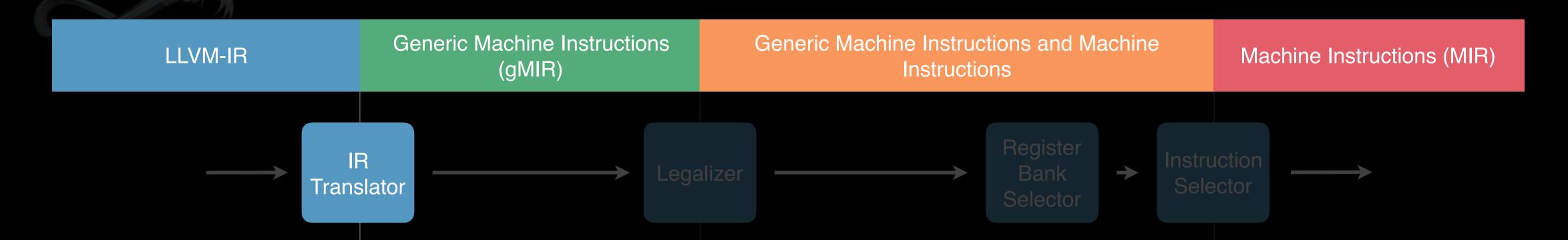
Quentin Colombet

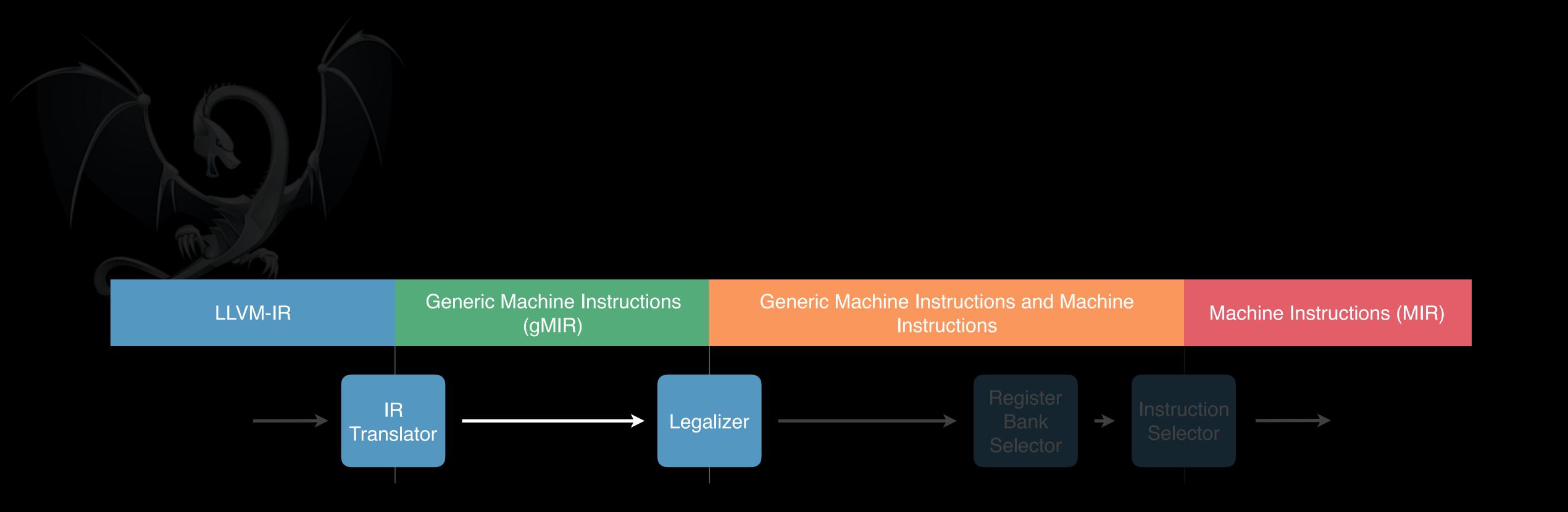
2015 LLVM Developers' Meeting



### IR Translator

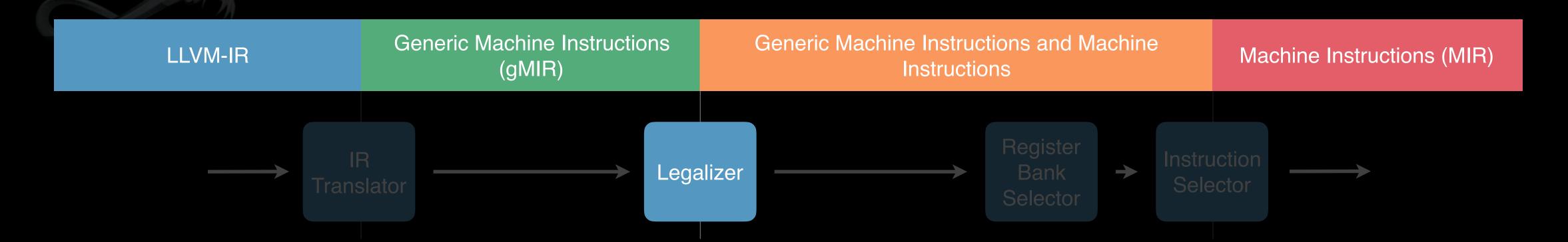
#### Convert LLVM-IR into gMIR

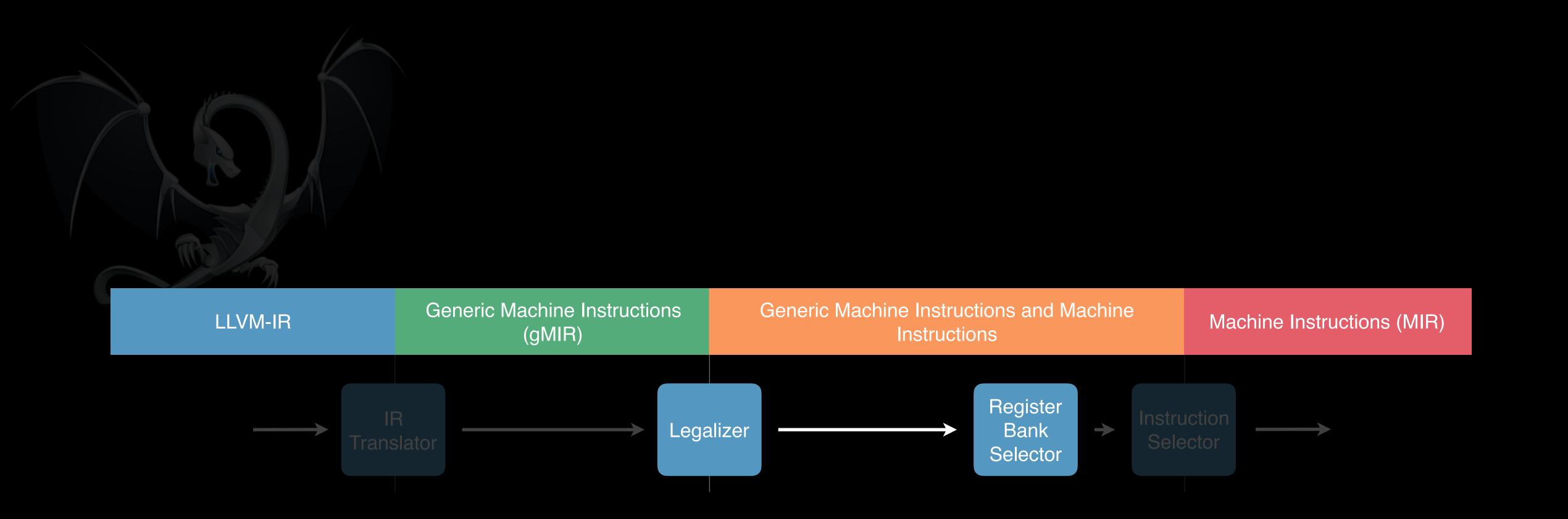




# Legalizer

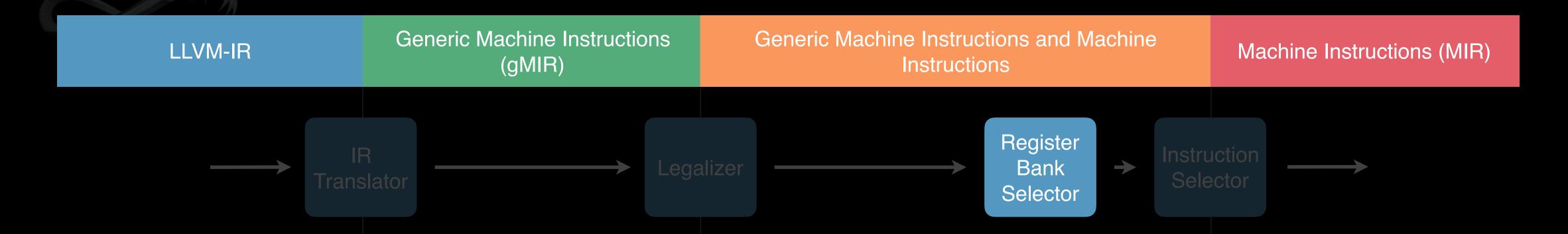
Replace unsupported operations with supported ones

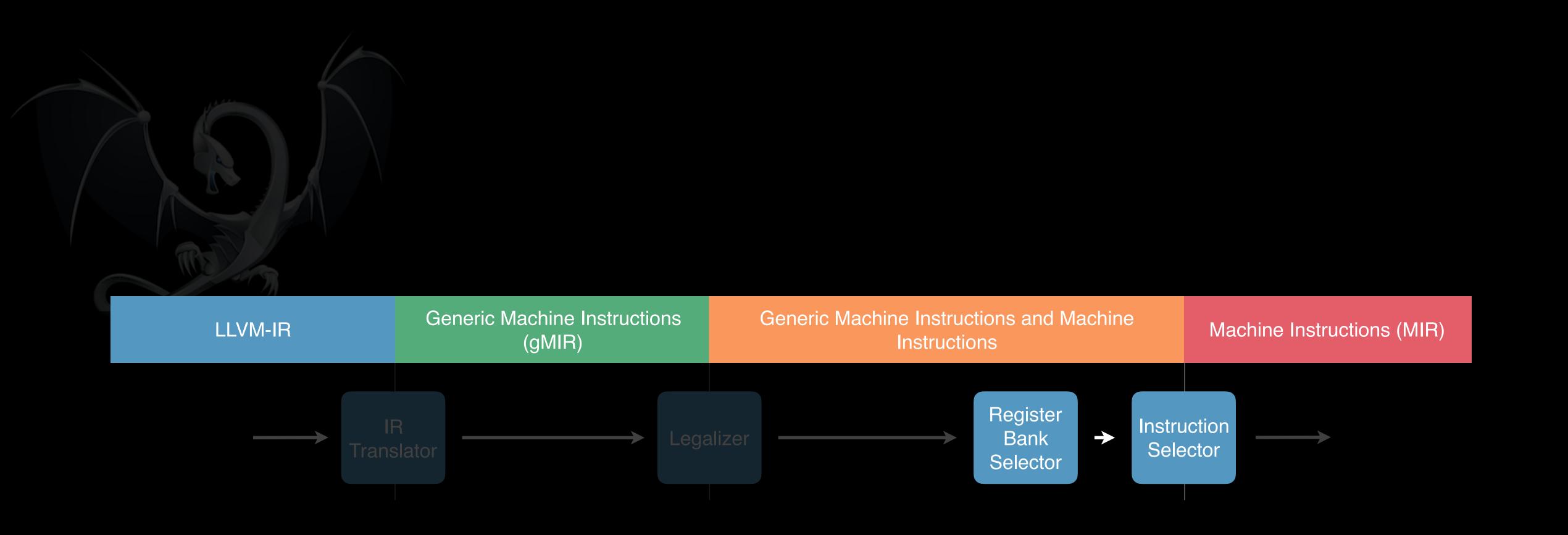




## Register Bank Selector

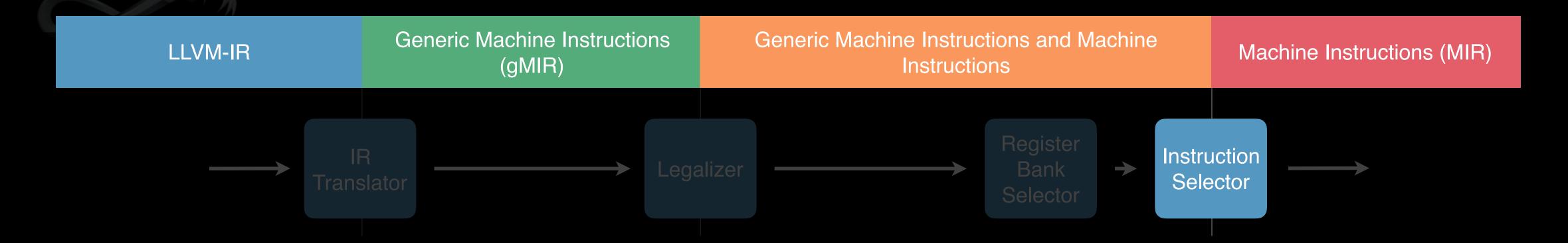
Binds registers to a Register Bank

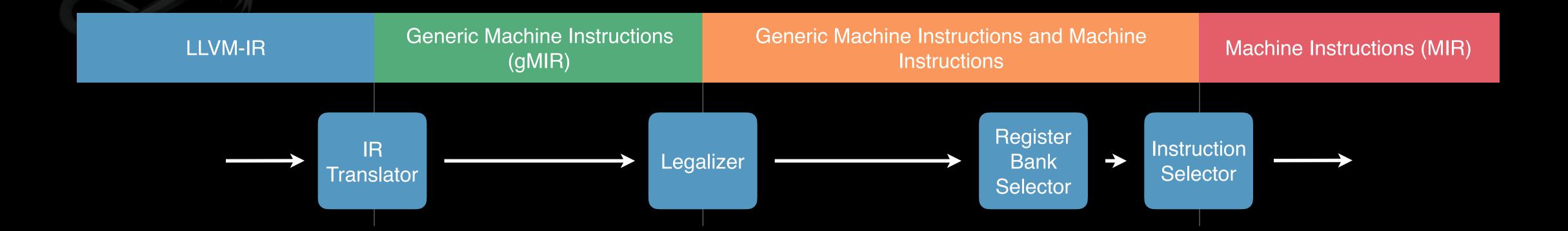


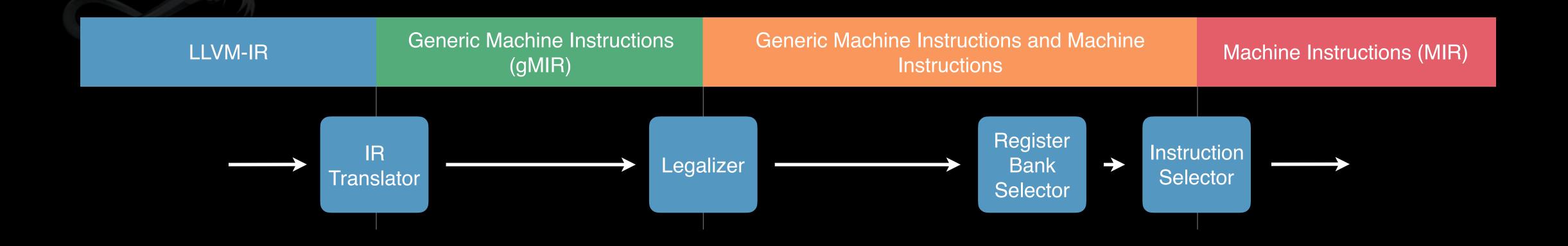


### Instruction Selector

Select target instructions

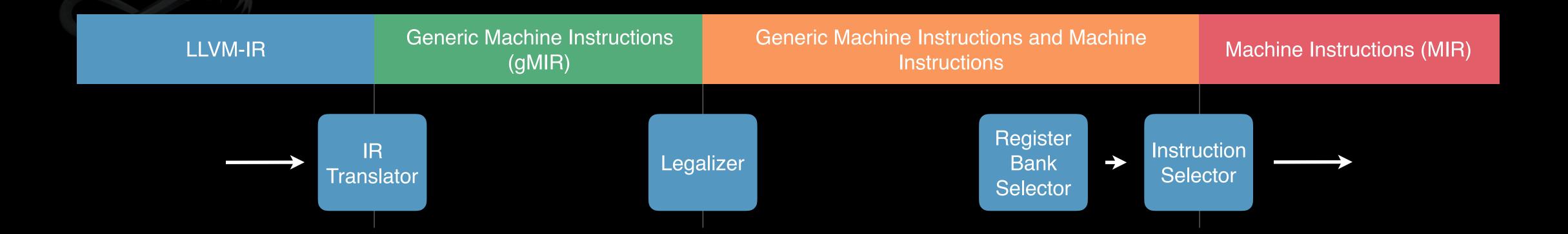




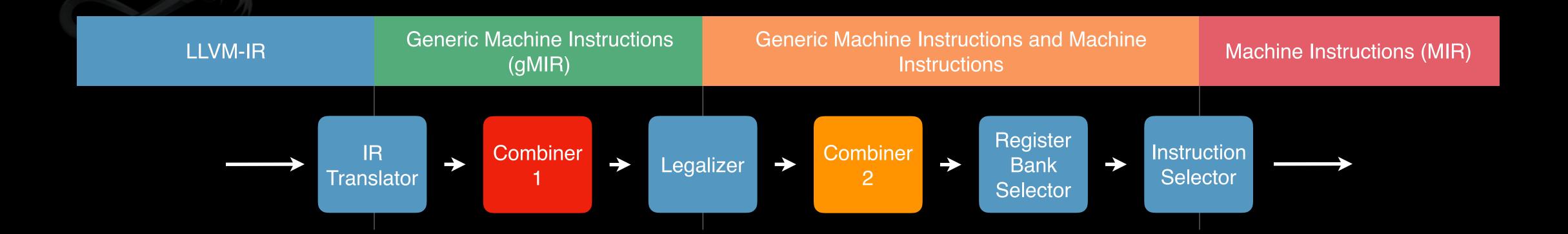


#### Tutorial: Head First into GloballSel

Aditya Nandakumar, Daniel Sanders, and Justin Bogner 2017 LLVM Developers' Meeting

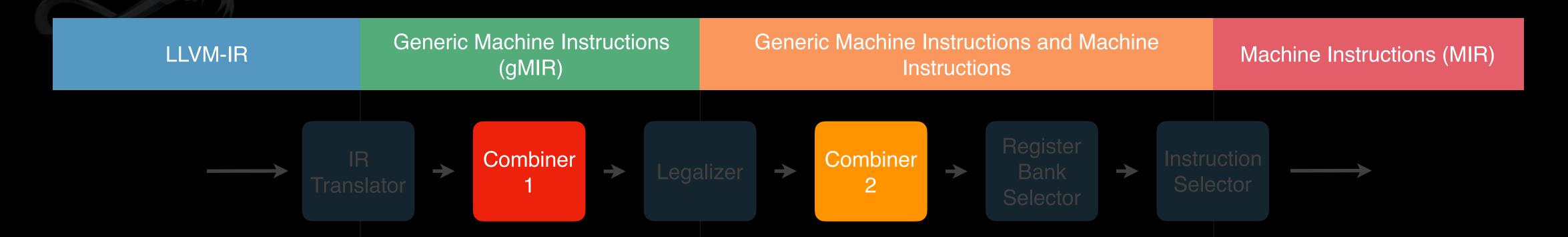


### Combiner



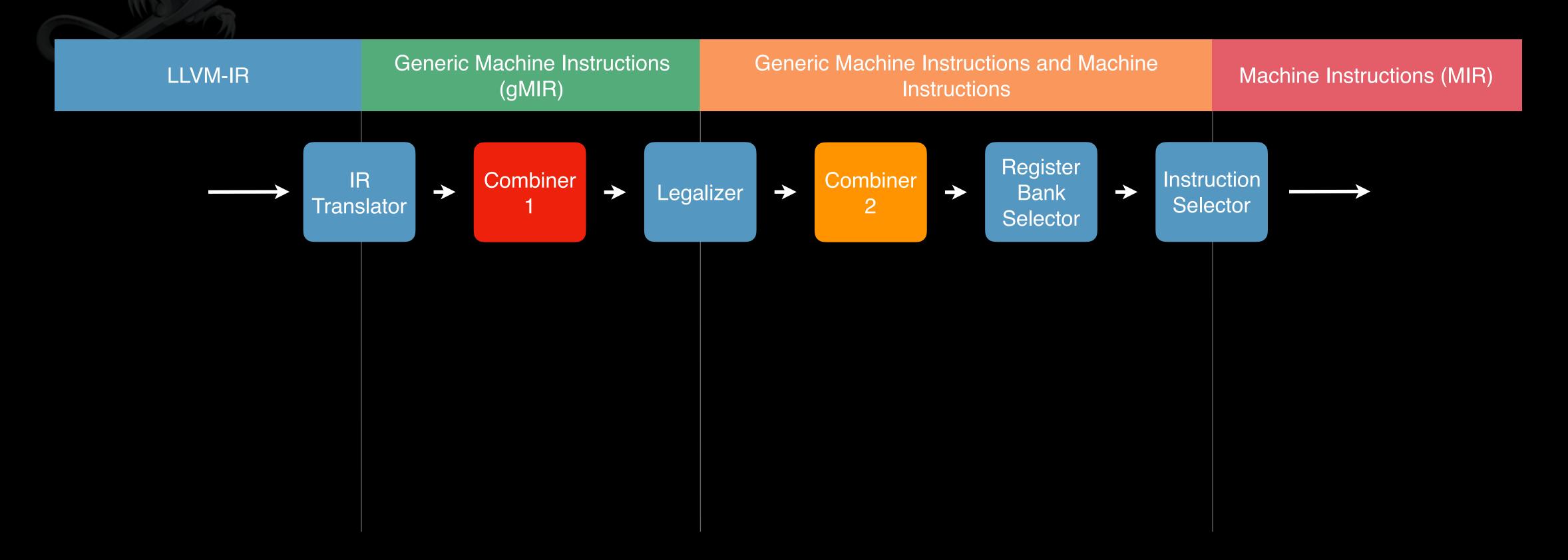


#### Simplify/Optimize gMIR/MIR



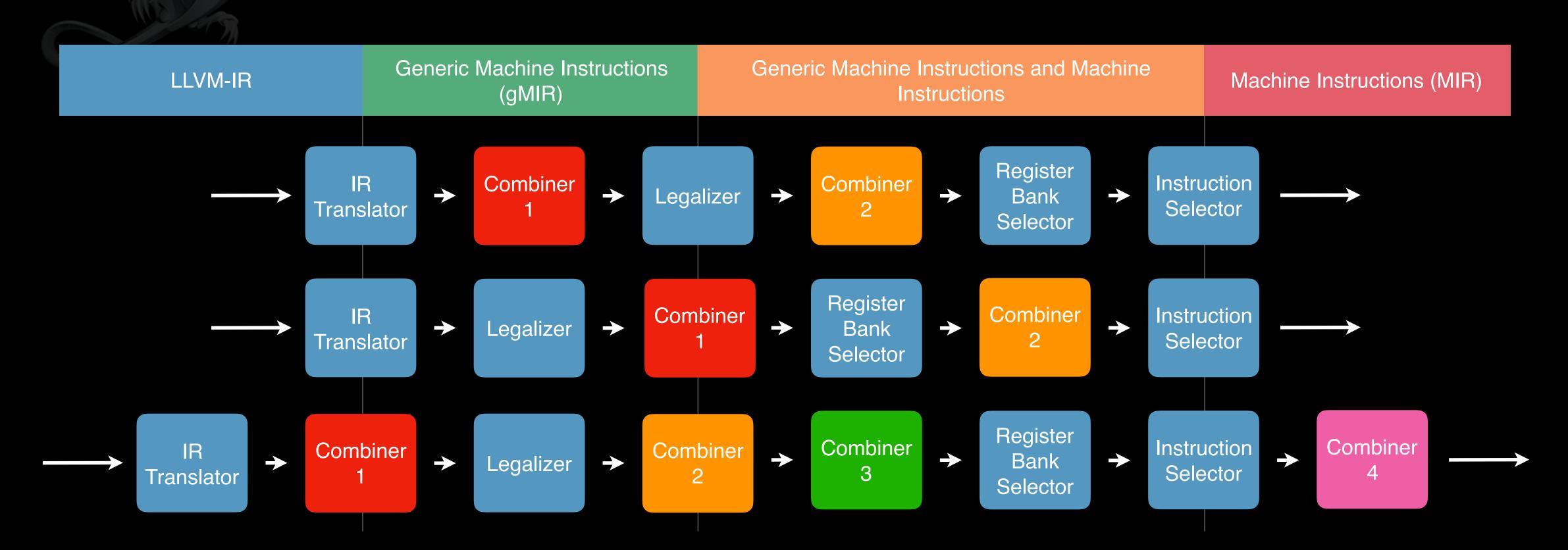
### Combiner

#### Simplify/Optimize gMIR/MIR



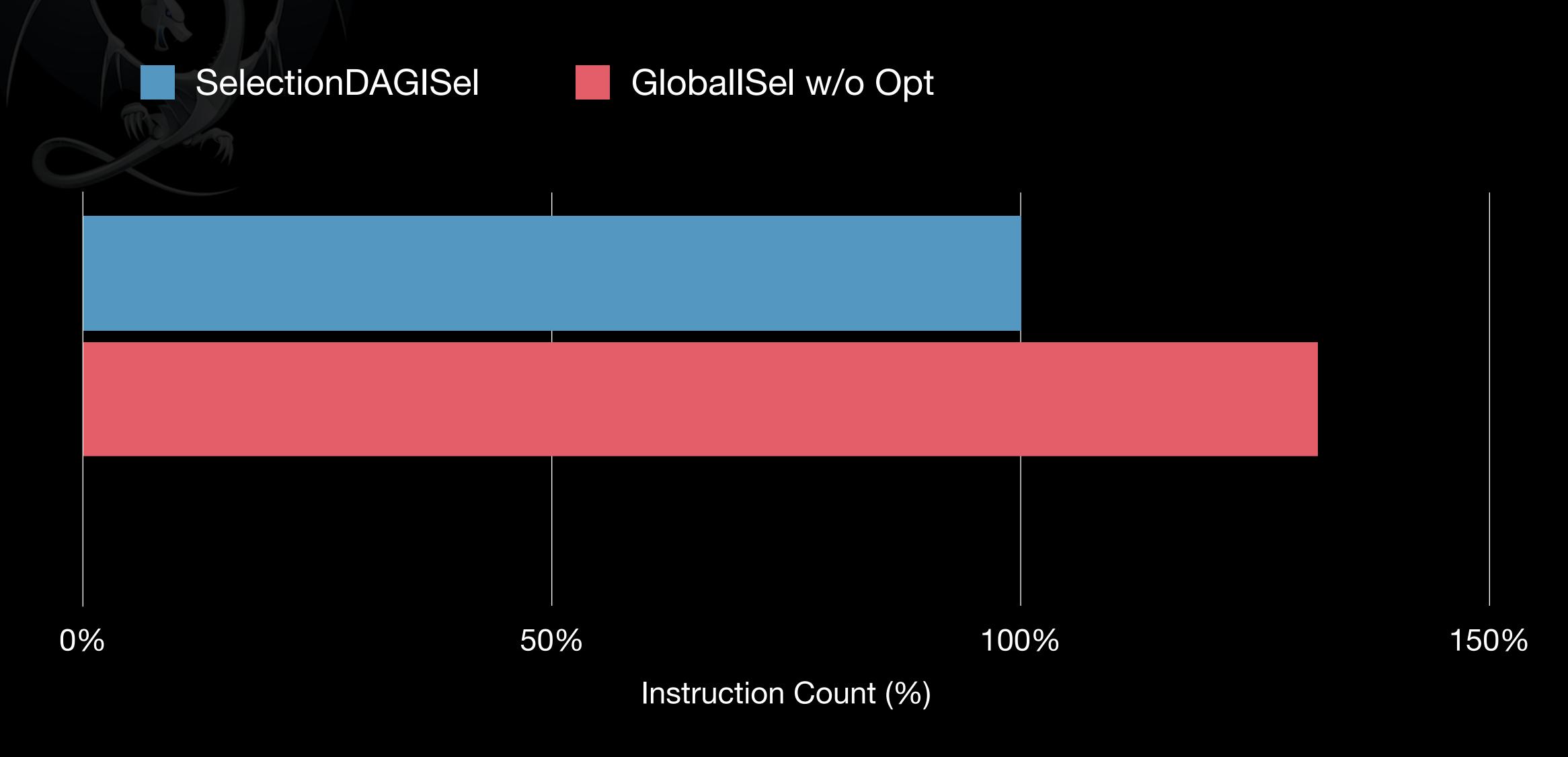
### Combiner

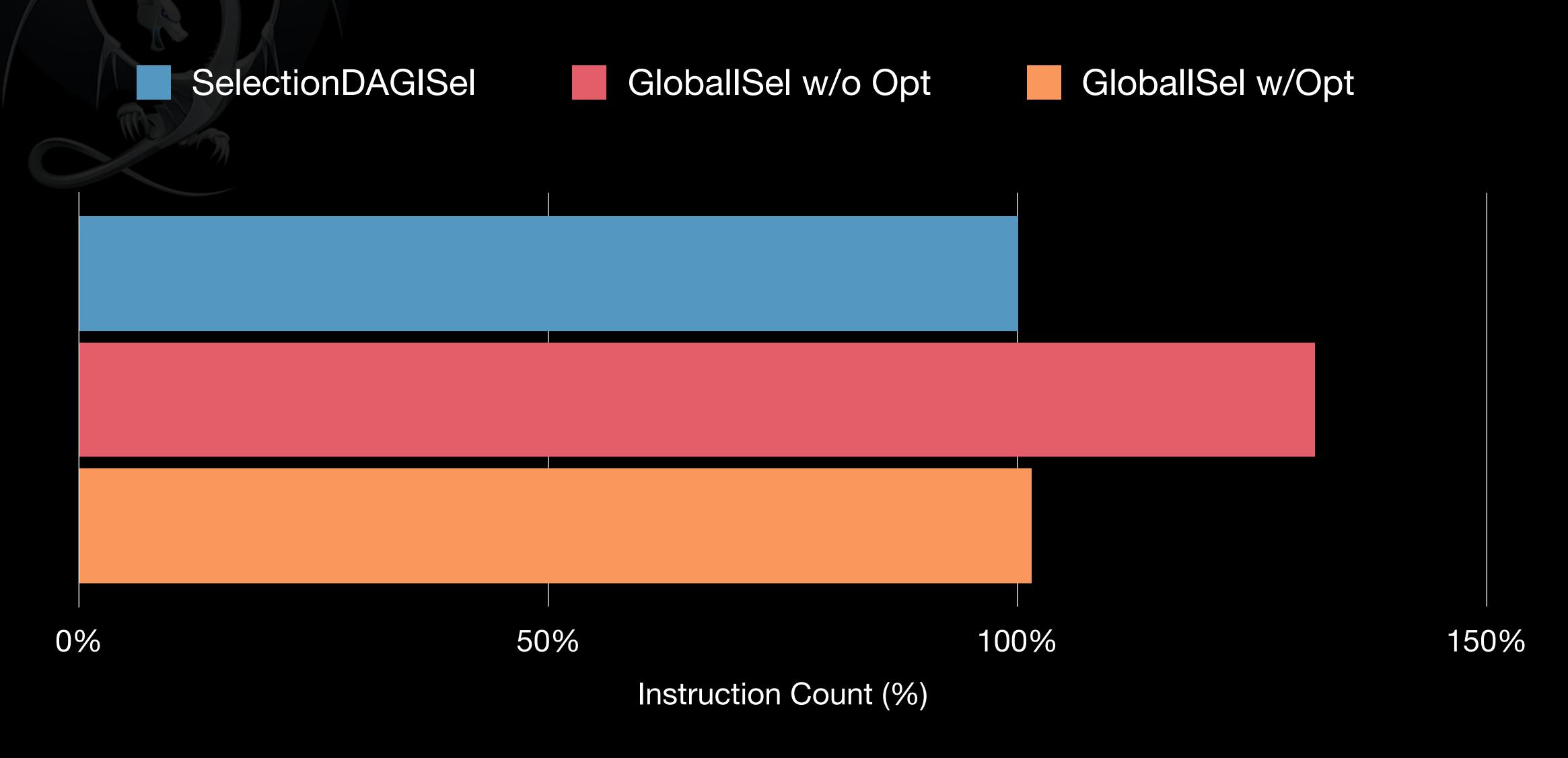
#### Simplify/Optimize gMIR/MIR

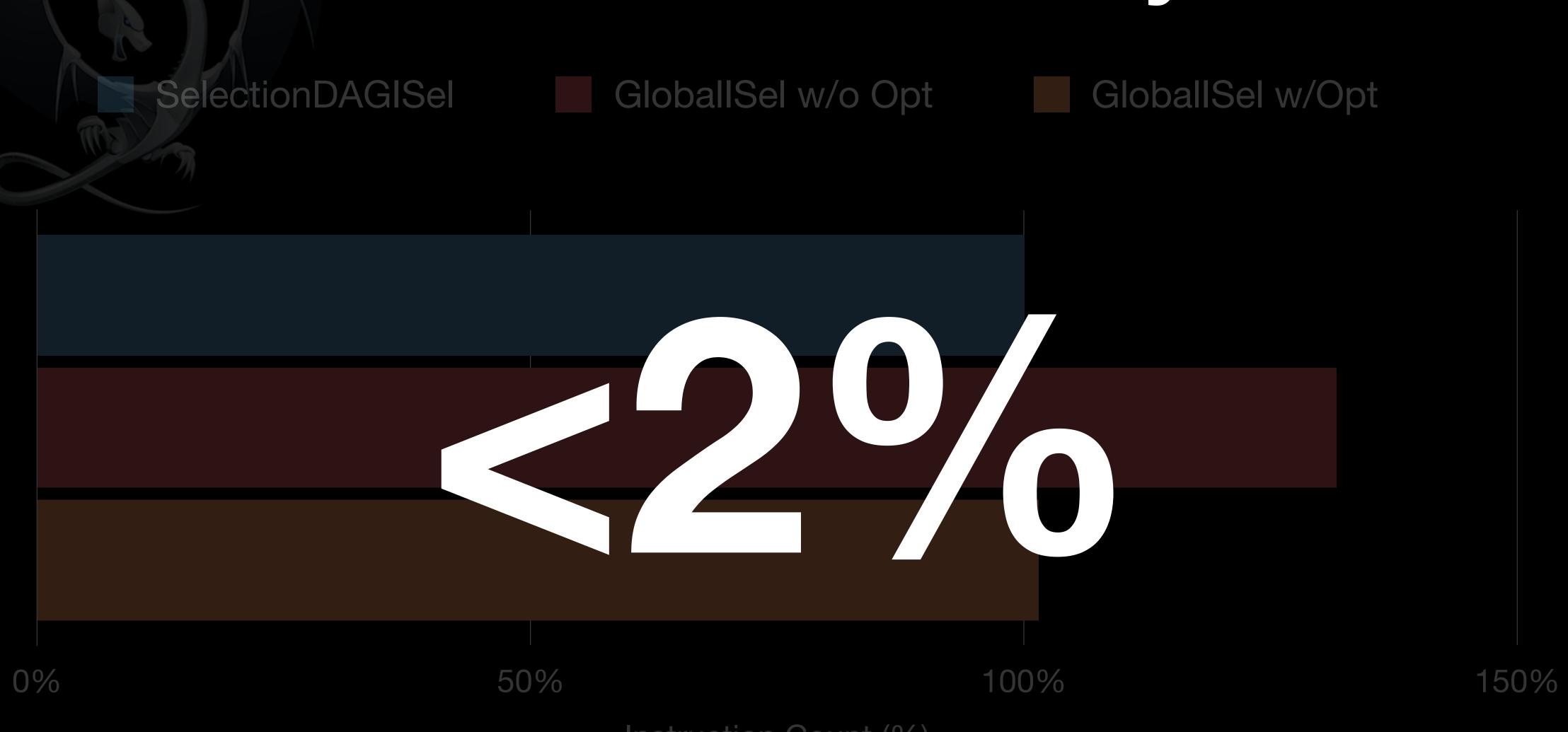




## Why do we need combiners?

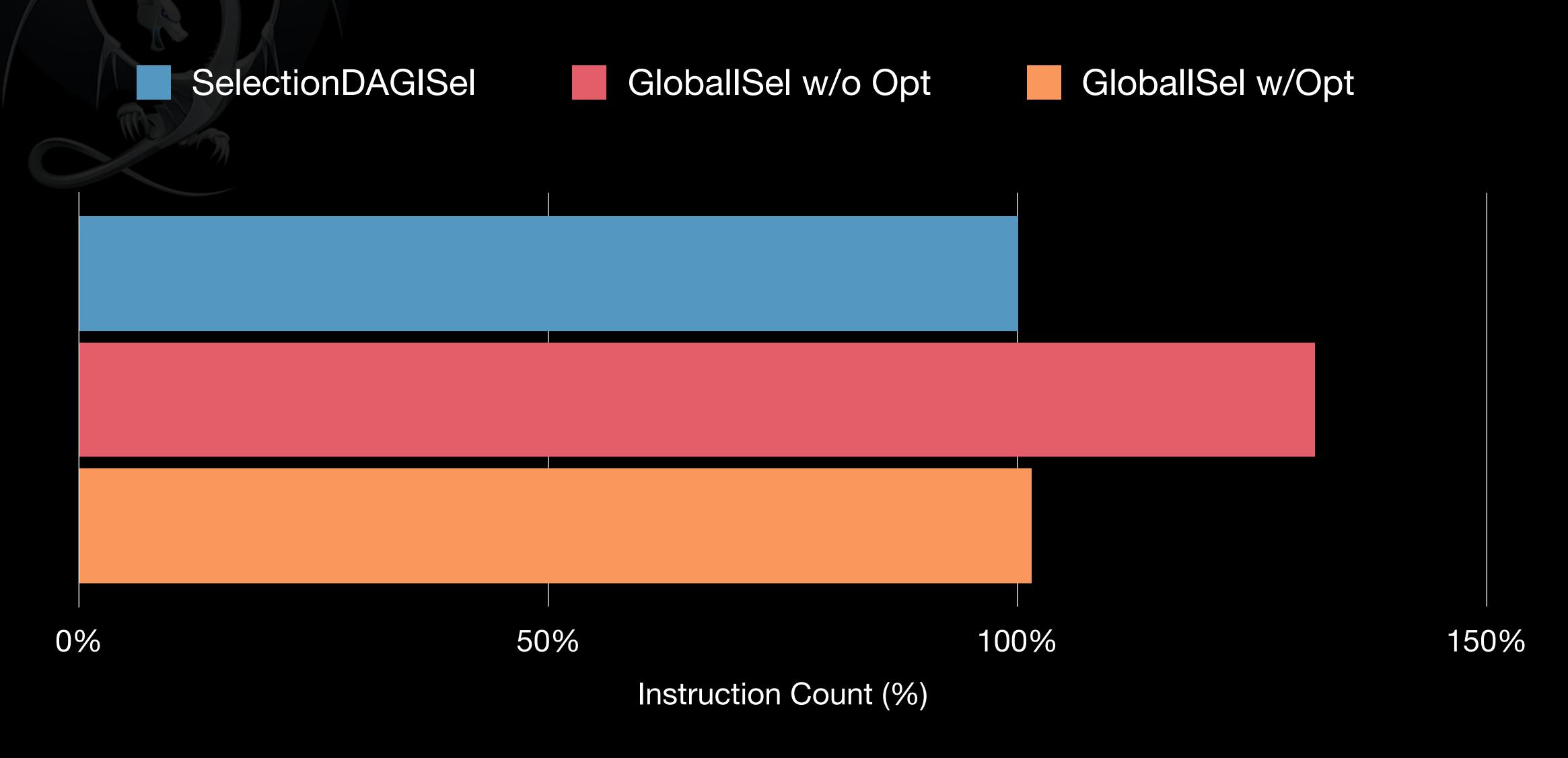




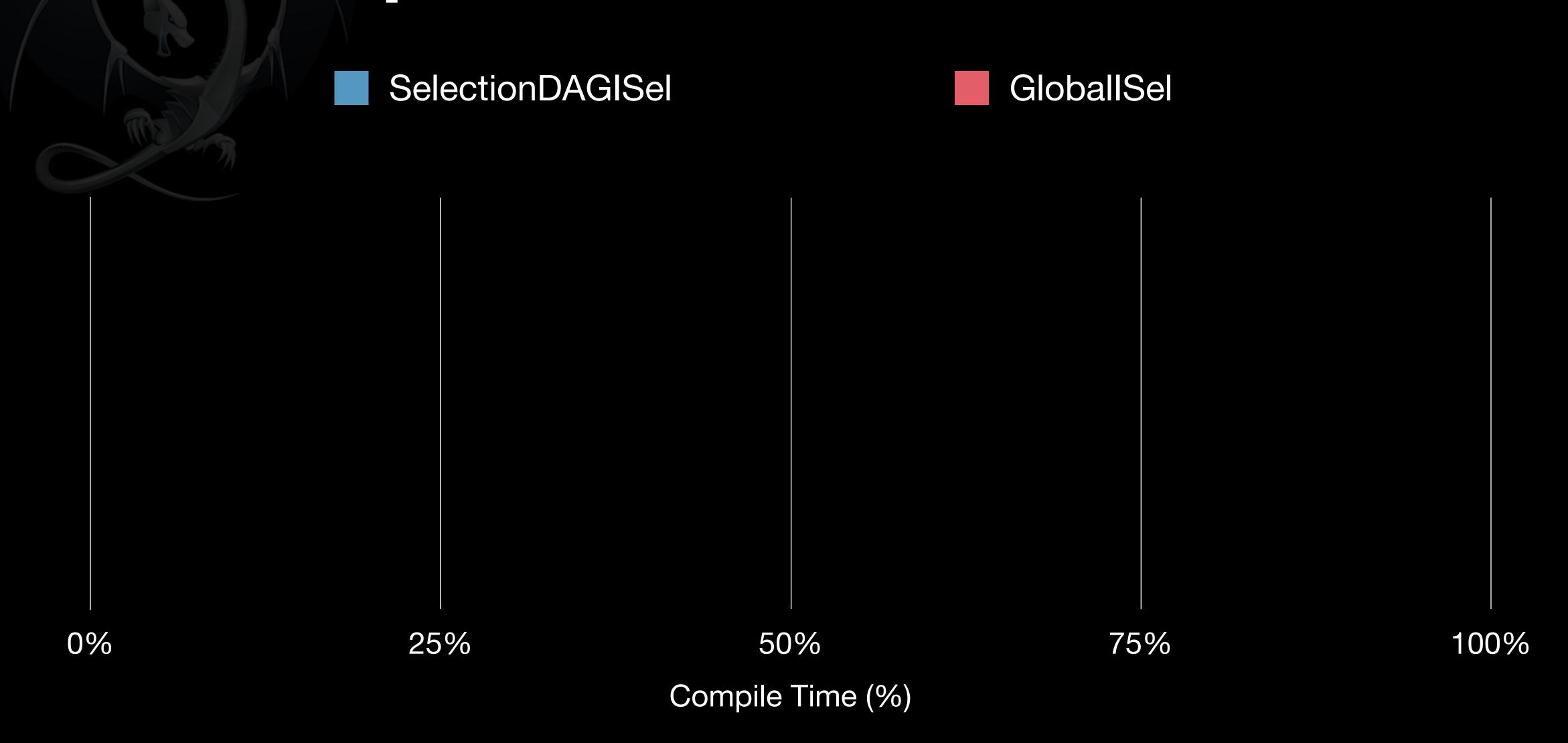


Instruction Count (%)

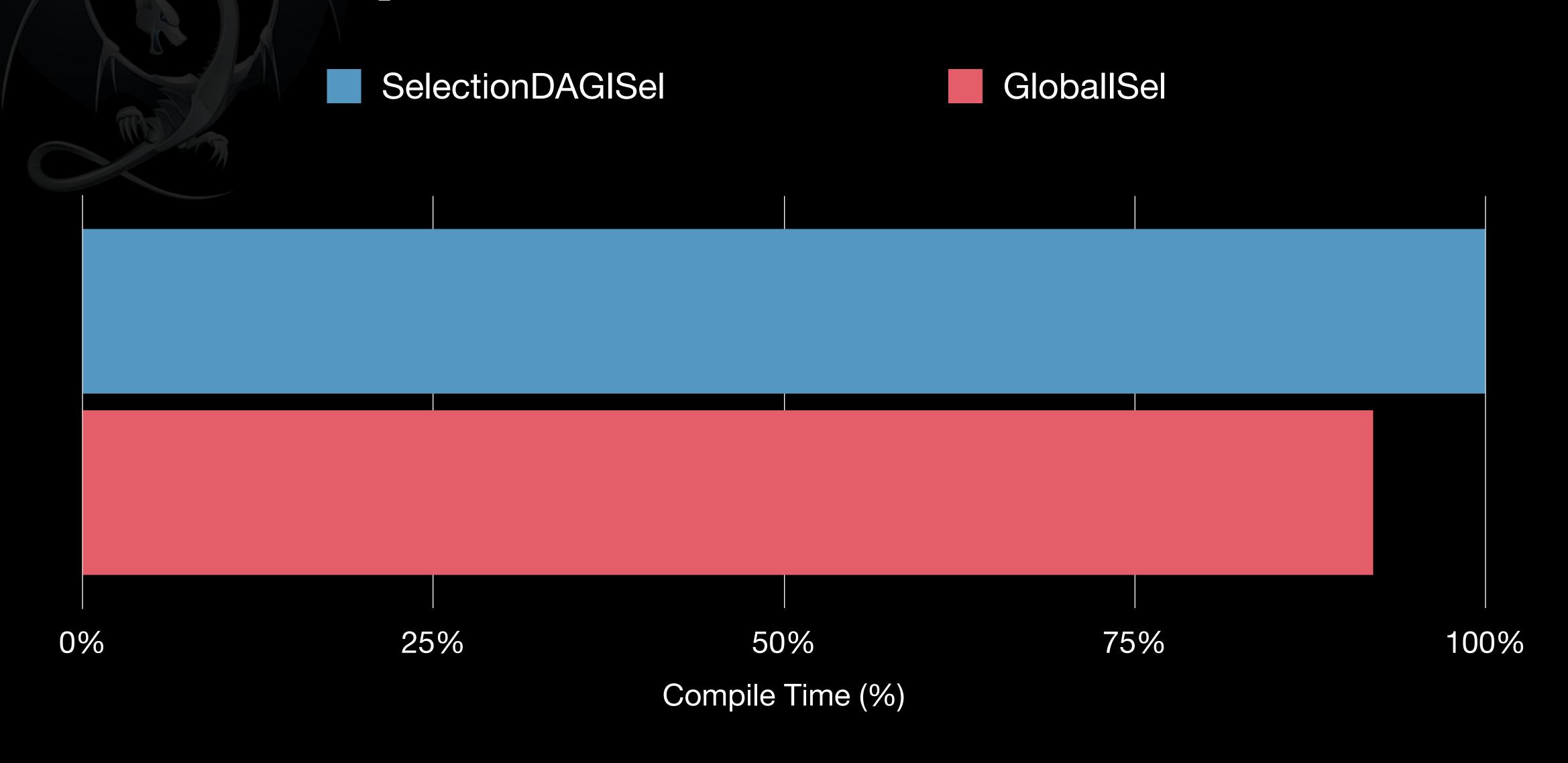
22



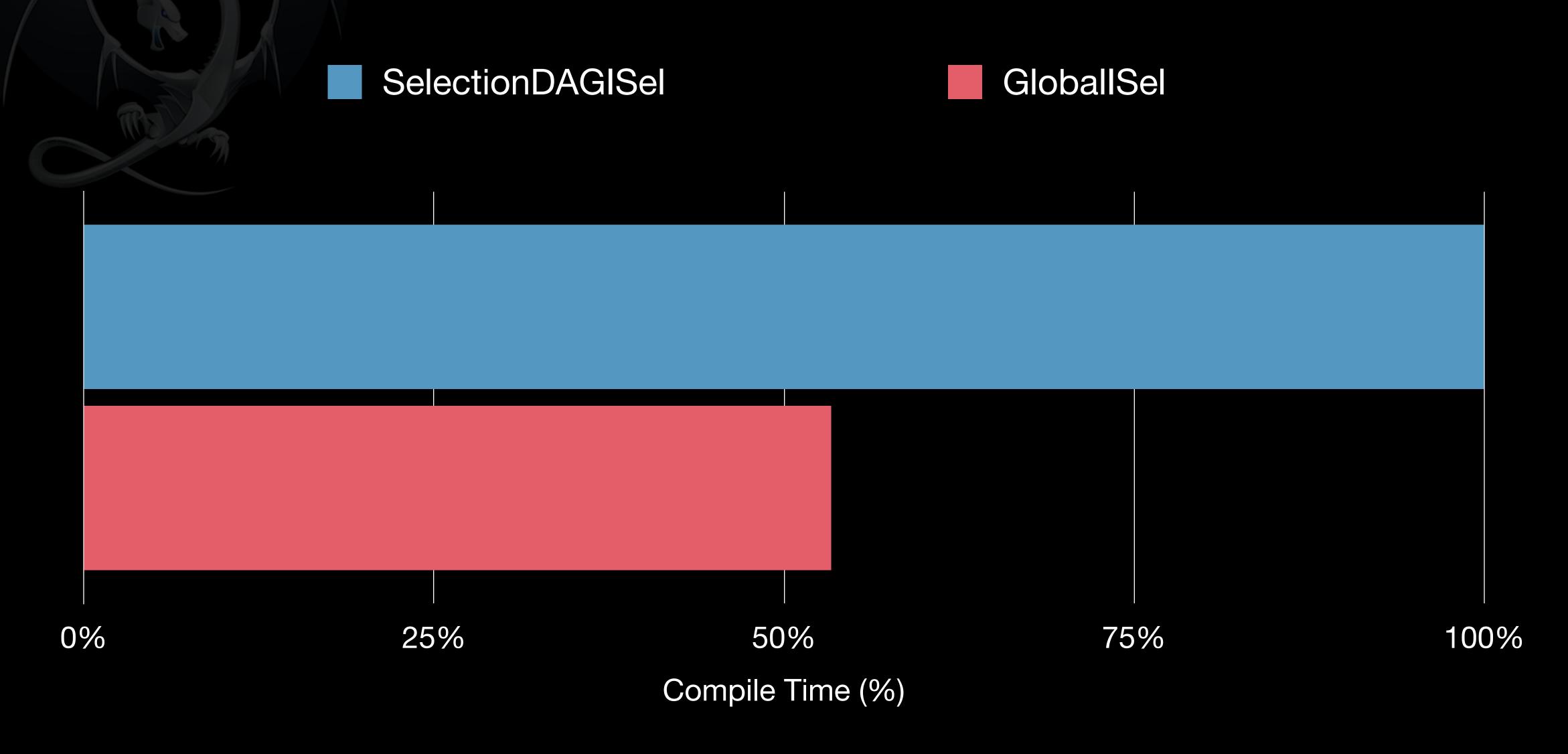
## Compile Time Performance



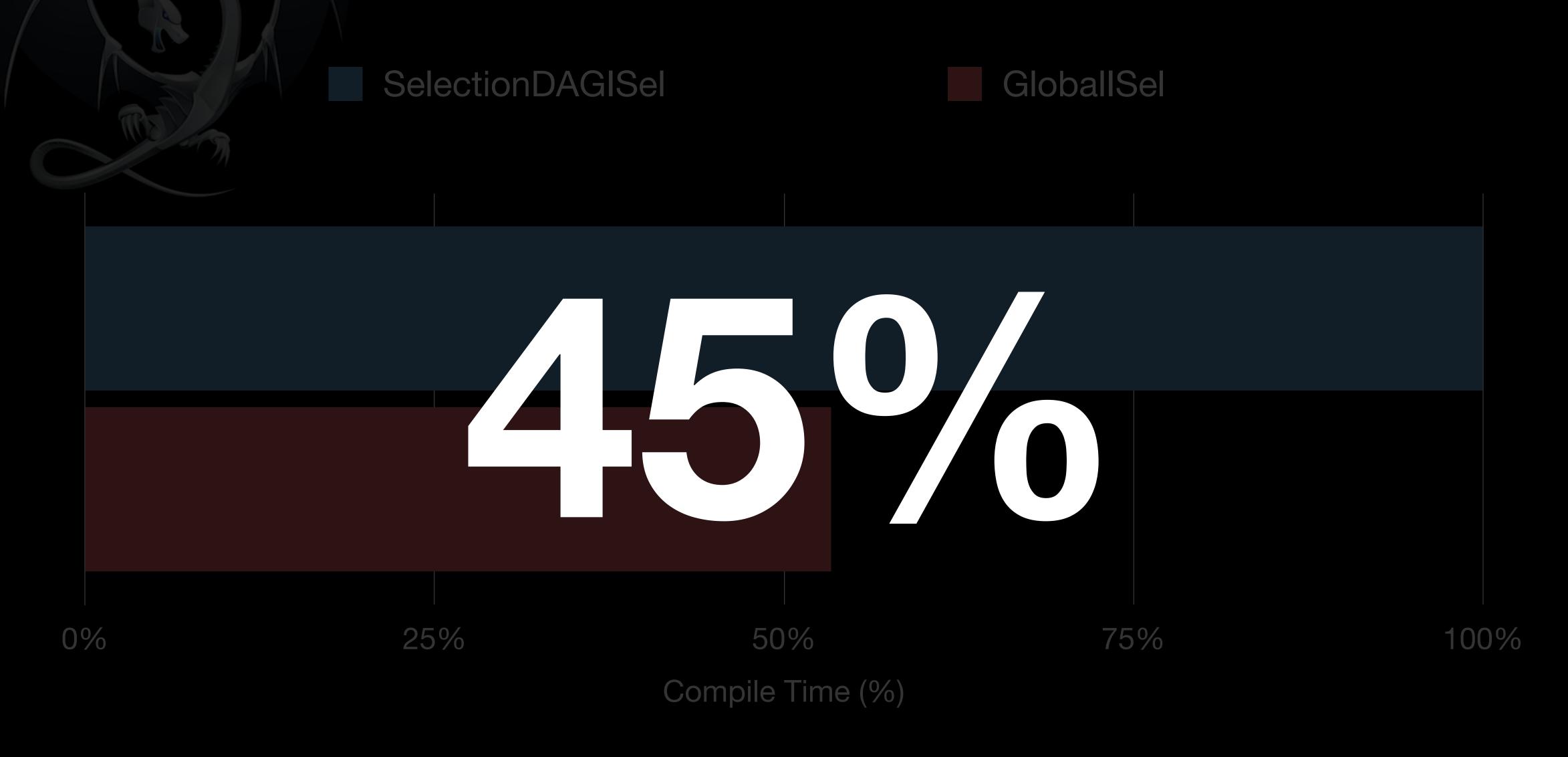
## Compile Time Performance



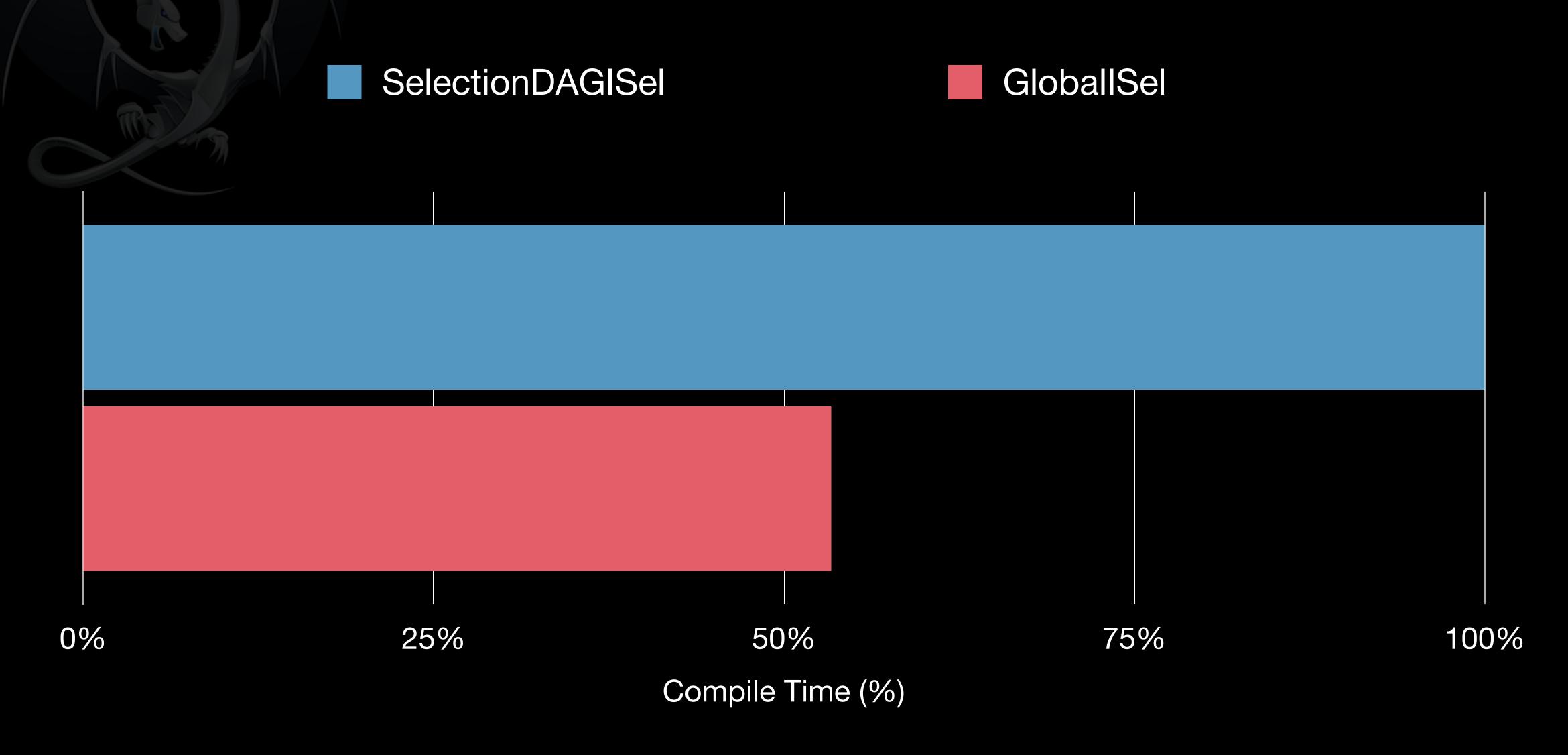
### Compile Time Performance - ISel Only



### Compile Time Performance - ISel Only



## Compile Time Performance - ISel Only





- Common Subexpression Elimination (CSE)
- Combiners
- KnownBits
- SimplifyDemandedBits

## CSE

Considered using MachineCSE, but it was expensive

## CSE

- Considered using MachineCSE, but it was expensive
- We chose a continuous CSE approach

#### CSE

- Considered using MachineCSE, but it was expensive
- We chose a continuous CSE approach
- Instructions are CSE'd at creation time using CSEMIRBuilder
  - Information is provided by an analysis pass
  - BasicBlock-local
  - Supports a subset of generic operations

## Things to be aware of

- CSE needs to be informed of:
  - Changes to Machinelnstrs (creation, modification, and erasure)
- Installs a delegate to handle creation/erasure automatically
- Installs a change observer to inform changes

# Compile Time Cost

- We were expecting this to come at a big compile-time cost
- Improved compile time for some cases
  - Later passes had less work to do



## Combiner

- Applies a set of combine rules
- Important for producing good code
- Expensive in terms of compile-time

```
define i32 @foo(i8 %in) {
  %ext1 = zext i8 %in to i16
  %ext2 = zext i16 %ext1 to i32
  ret i32 %ext2
}
```

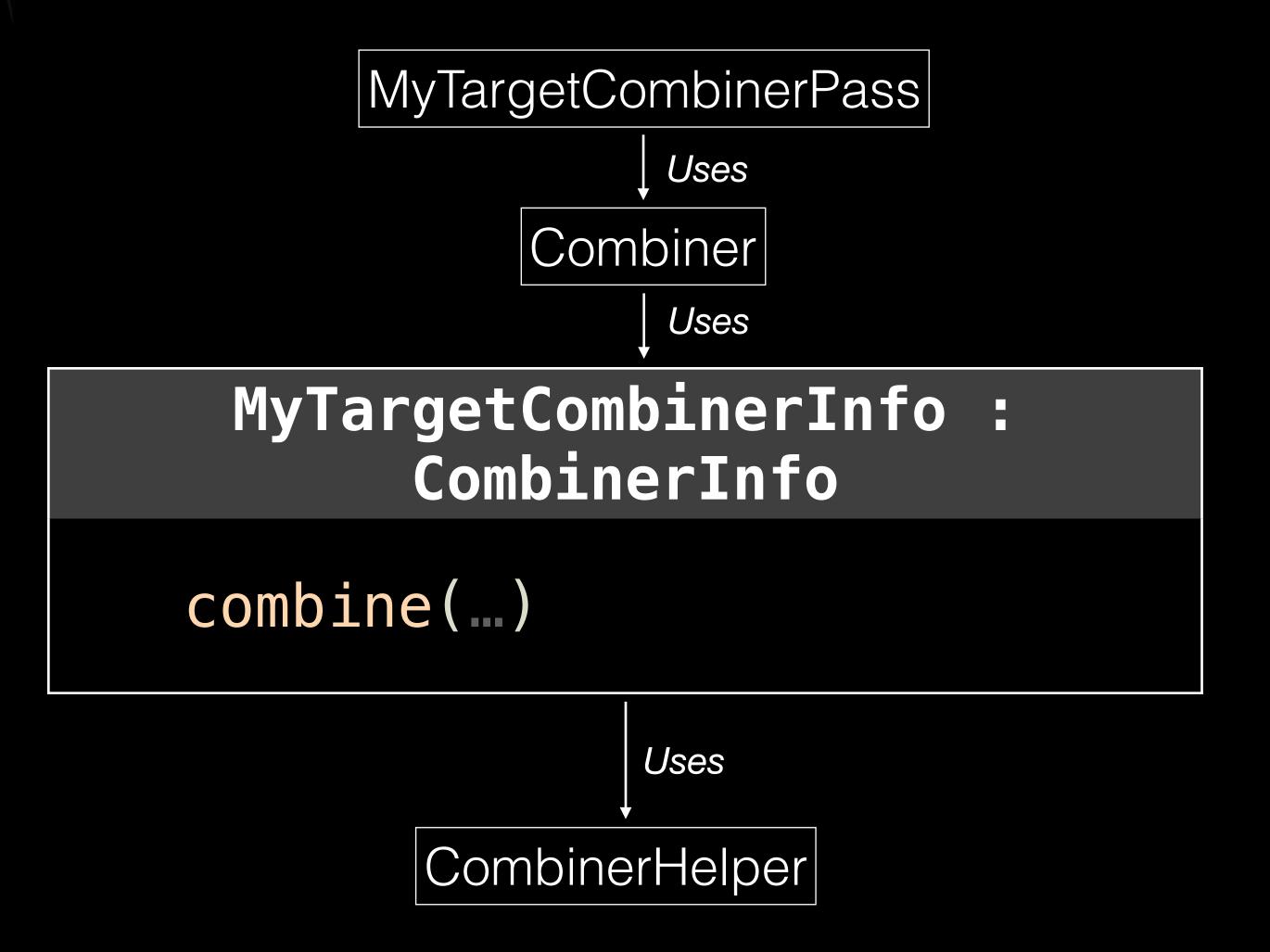
```
define i32 @foo(i8 %in) {
  %ext1 = zext i8 %in to i16
  %ext2 = zext i16 %ext1 to i32
  ret i32 %ext2
}
```

```
define i32 @foo(i8 %in) {
   %ext2 = zext i8 %in to i32
   ret i32 %ext2
}
```

## Globall Sel Combiner

- GloballSel Combiner consists of 3 main pieces
  - Combiner iterates over the MachineFunction
  - CombinerInfo specifies which operations to be combined and how
  - CombinerHelper is a library of generic combines

## GloballSelCombiner



#### A Basic Combiner

```
bool MyTargetCombinerInfo::combine(GISelChangeObserver &Observer,
                                   MachineInstr &MI,
                                   MachineIRBuilder &B) const {
    MyTargetCombinerHelper TCH(Observer, B, KB);
    // Try all combines.
    if (OptimizeAggresively)
        return TCH.tryCombine(MI);
    // Combine COPY only.
    if (MI.getOpcode() == TargetOpcode::COPY)
        return TCH.tryCombineCopy(MI);
    return false;
```

## A Simple Combine

```
bool MyTargetCombinerHelper::combineExt(GISelChangeObserver &Observer,
       MachineInstr &MI, MachineIRBuilder &B) const {
    // Combine zext(zext x) -> zext x
    if (MI.getOpcode() == TargetOpcode::G_ZEXT) {
        Register SrcReg = MI.getOperand(1).getReg();
        MachineInstr *SrcMI = MRI.getVRegDef(SrcReg);
        // Check if SrcMI is a G_ZEXT.
        if (SrcMI->getOpcode() == TargetOpcode::G_ZEXT) {
            SrcReg = SrcMI->getOperand(1).getReg();
            B.buildZExt(Reg, SrcReg);
            MI.eraseFromParent();
            return true;
```

## A Simple Combine

```
bool MyTargetCombinerHelper::combineExt(GISelChangeObserver &Observer,
       MachineInstr &MI, MachineIRBuilder &B) const {
      Combine zext(zext x) -> zext x
    if (MI.getOpcode() == TargetOpcode::G_ZEXT) {
        Register SrcReg = MI.getOperand(1).getReg();
        MachineInstr *SrcMI = MRI.getVRegDef(SrcReg);
        // Check if SrcMI is a G_ZEXT.
        if (SrcMI->getOpcode() == TargetOpcode::G_ZEXT) {
            SrcReg = SrcMI->getOperand(1).getReg();
            B.buildZExt(Reg, SrcReg);
            MI.eraseFromParent();
            return true;
```

## MIPatternMatch

- Simple and easy mechanism to match generic patterns
- Similar to what we have for LLVM IR
- Combines can be implemented easily using matchers

#### MIPatternMatch

- Simple and easy mechanism to match generic patterns
- Similar to what we have for LLVM IR
- Combines can be implemented easily using matchers

```
// Combine zext(zext x) -> zext x
Register SrcReg;
if (mi_match(Reg, MRI, m_GZext(m_GZext(m_Reg(SrcReg))))) {
   B.buildZExt(Reg, SrcReg);
   MI.eraseFromParent();
   return true;
}
```

## A Simpler Combine

```
// Combine zext(zext x) -> zext x
Register SrcReg;
if (mi_match(Reg, MRI, m_GZext(m_GZext(m_Reg(SrcReg))))) {
   B.buildZExt(Reg, SrcReg);
   MI.eraseFromParent();
   return true;
}
```

## A Simpler Combine

```
// Combine zext(zext x) -> zext x
Register SrcReg;
if (mi_match(Reg, MRI, m_GZext(m_GZext(m_Reg(SrcReg))))) {
  B.buildZExt(Reg, SrcReg);
 MI.eraseFromParent();
  return true;
// Combine zext(zext x) -> zext x
Register SrcReg;
if (mi_match(Reg, MRI, m_GZext(m_GZext(m_Reg(SrcReg))))) {
  Observer.changingInstr(MI);
 MI.getOperand(1).setReg(SrcReg);
  Observer.changedInstr(MI);
  return true;
```

## A Simpler Combine

```
// Combine zext(zext x) -> zext x
Register SrcReg;
if (mi_match(Reg, MRI, m_GZext(m_GZext(m_Reg(SrcReg))))) {
  B.buildZExt(Reg, SrcReg);
 MI.eraseFromParent();
  return true;
// Combine zext(zext x) -> zext x
Register SrcReg;
if (mi_match(Reg, MRI, m_GZext(m_GZext(m_Reg(SrcReg))))) {
  Observer.changingInstr(MI);
 MI.getOperand(1).setReg(SrcReg);
  Observer.changedInstr(MI);
  return true;
```

# Informing the Observer

- Observer needs to be informed when something changed
  - createdInstr() and erasedInstr() are handled automatically
  - changingInstr() and changedInstr() are handled manually and mandatory for MRI.setRegClass(), MO.setReg(), etc.

Many combines are only valid for certain cases

- Many combines are only valid for certain cases
  - ►  $(a + 1) \rightarrow (a | 1)$  is only valid if (a & 1) == 0

- Many combines are only valid for certain cases
  - ►  $(a + 1) \rightarrow (a | 1)$  is only valid if (a & 1) == 0
- We added an analysis pass to provide this information

- Many combines are only valid for certain cases
  - ► (a + 1)  $\rightarrow$  (a | 1) is only valid if (a & 1) == 0
- We added an analysis pass to provide this information
- Currently provides known-ones, known-zeros, and unknowns

```
%1:(s32) = G_CONSTANT i32 0xFF0
%2:(s32) = G_AND %0, %1
%3:(s32) = G_CONSTANT i32 0x0FF
%4:(s32) = G_AND %2, %3
```

	Value
%0	0x???????
%1	0x00000FF0
%2	
%3	0x000000FF
%4	

```
%1:(s32) = G_CONSTANT i32 0xFF0
%2:(s32) = G_AND %0, %1
%3:(s32) = G_CONSTANT i32 0x0FF
%4:(s32) = G_AND %2, %3
```

	Value
%0	0x???????
%1	0x00000FF0
%2	0x00000??0
%3	0x000000FF
%4	0x000000?0

```
%1:(s32) = G_CONSTANT i32 0xFF0
%2:(s32) = G_AND %0, %1
%3:(s32) = G_CONSTANT i32 0x0FF
```

 $%4:(s32) = G_AND %2, %3$ 

	Value
%0	0x???????
%1	0x00000FF0
%2	0x00000??0
%3	0x000000FF
%4	0x000000?0



 $%5:(s32) = G_CONSTANT i32 0x0F0$ 

 $%4:(s32) = G_AND %2, %3$ 

	Value
<b>%0</b>	0x???????
%1	0x00000FF0
<del>2</del>	0x00000??0
%3	0x00000FF
%4	0x000000?0
%5	0x00000F0

# Why an Analysis Pass?

- In SelectionDAGISel, computeKnownBits() is just a function
- In GloballSel, it's an Analysis Pass

# Why an Analysis Pass?

- In SelectionDAGISel, computeKnownBits() is just a function
- In GloballSel, it's an Analysis Pass
- It allows us to add support for:
  - Caching within a pass
  - Caching between passes
  - Early exit when enough is known

# Why an Analysis Pass?

- In SelectionDAGISel, computeKnownBits() is just a function
- In GloballSel, it's an Analysis Pass
- It allows us to add support for:
  - Caching within a pass
  - Caching between passes
  - Early exit when enough is known
- Allows us to have alternative implementations

# Extending KnownBits

```
void MyTargetLowering::computeKnownBitsForTargetInstr(
         GISelKnownBits &Analysis, Register R, KnownBits &Known,
         const APInt &DemandedElts, const MachineRegisterInfo &MRI,
         unsigned Depth = 0) const override {
  switch (Opcode) {
  case TargetOpcode::ANDWrr: {
    Analysis.computeKnownBitsImpl(MI.getOperand(2).getReg(), Known, DemandedElts, Depth + 1);
    Analysis.computeKnownBitsImpl(MI.getOperand(1).getReg(), Known2, DemandedElts, Depth + 1);
    Known One &= Known One;
    Known Zero = Known Zero;
    break;
```

# KnownBits Analysis

- Allows optimizations that otherwise wouldn't be possible
- Available to any MachineFunction pass
- Caching will make it cheaper than SelectionDAGISel's equivalent

## SimplifyDemandedBits

- Essentially a special case of Combine
- Tries to eliminate calculations that contribute to the bits that are never read

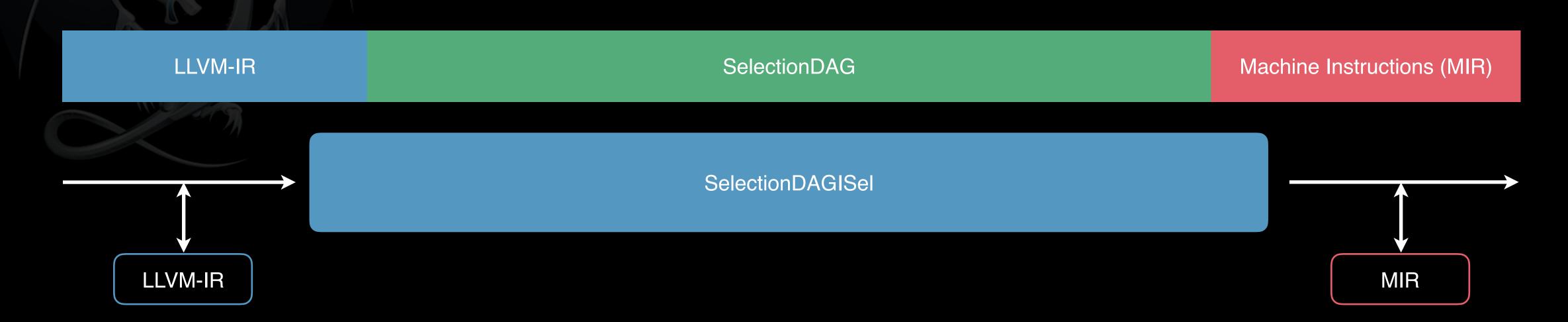
# SimplifyDemandedBits

- Essentially a special case of Combine
- Tries to eliminate calculations that contribute to the bits that are never read
- If demand mask is 0xF0:
  - ► (a << 16) (b & 0xFFFF)  $\rightarrow$  (b & 0xFFFF)

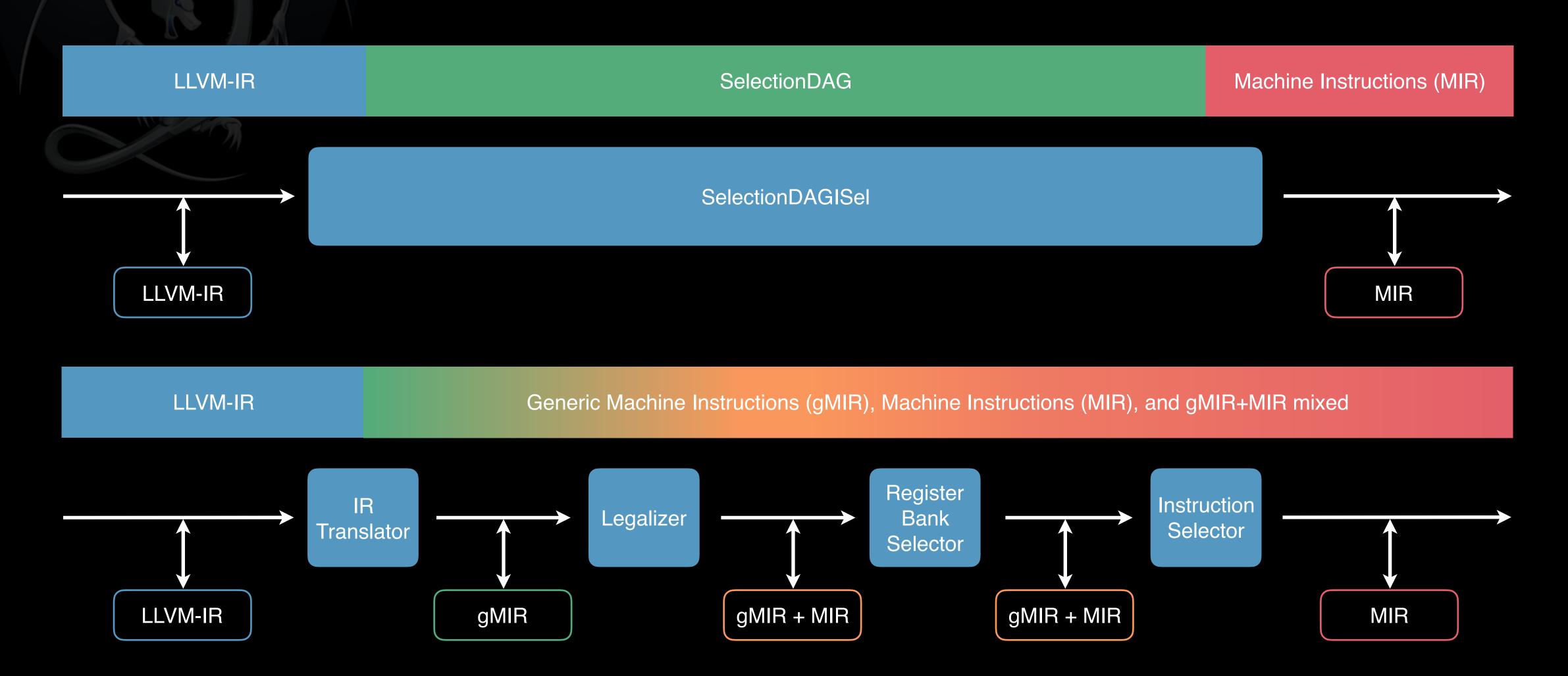
# SimplifyDemandedBits

- Essentially a special case of Combine
- Tries to eliminate calculations that contribute to the bits that are never read
- If demand mask is 0xF0:
  - ► (a << 16) | (b & 0xFFFF)  $\rightarrow$  (b & 0xFFFF)
- Not upstreamed yet, but we plan to fix that soon

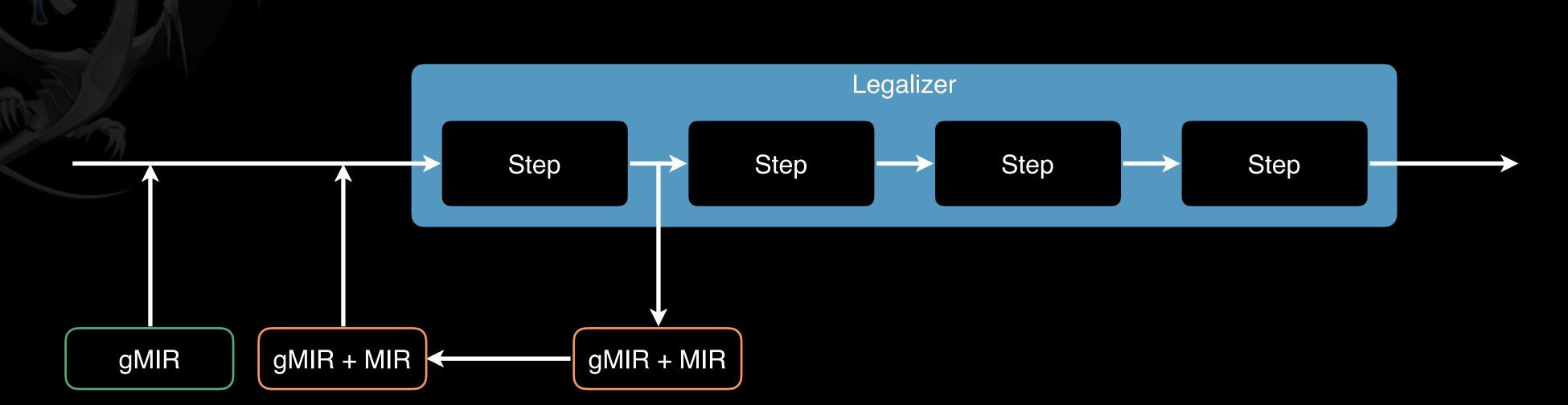
# Testing



## Testing



# Unit Testing

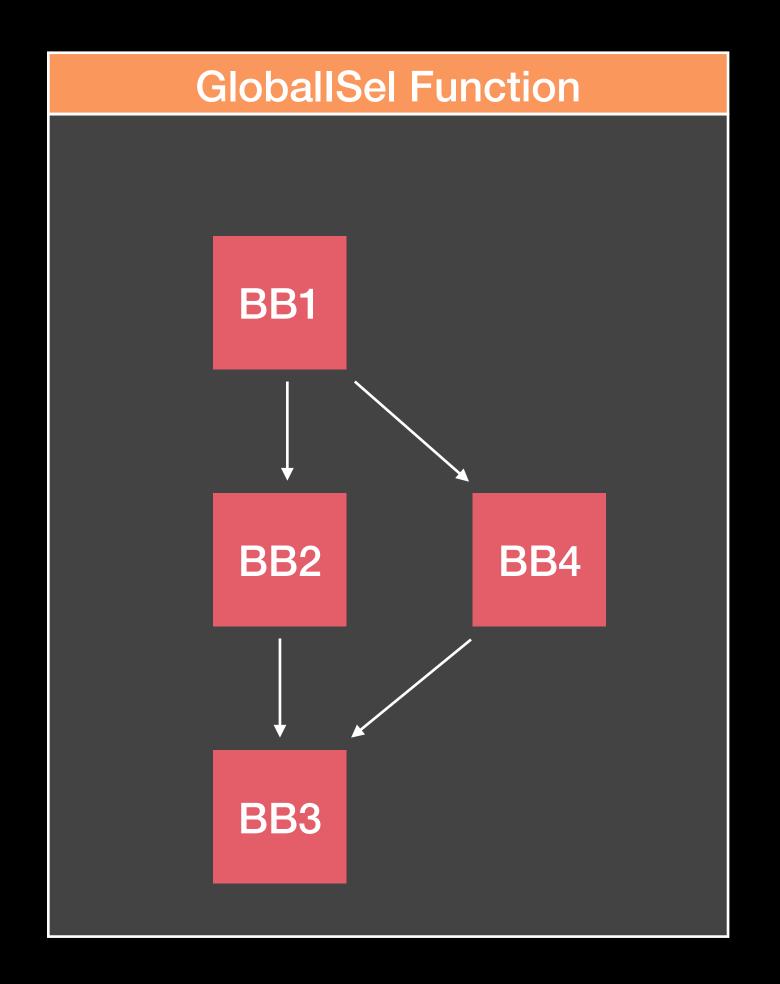


- Unit Testable too
  - We use FileCheck as a library to check results
  - It allows us to test exactly what optimizations do

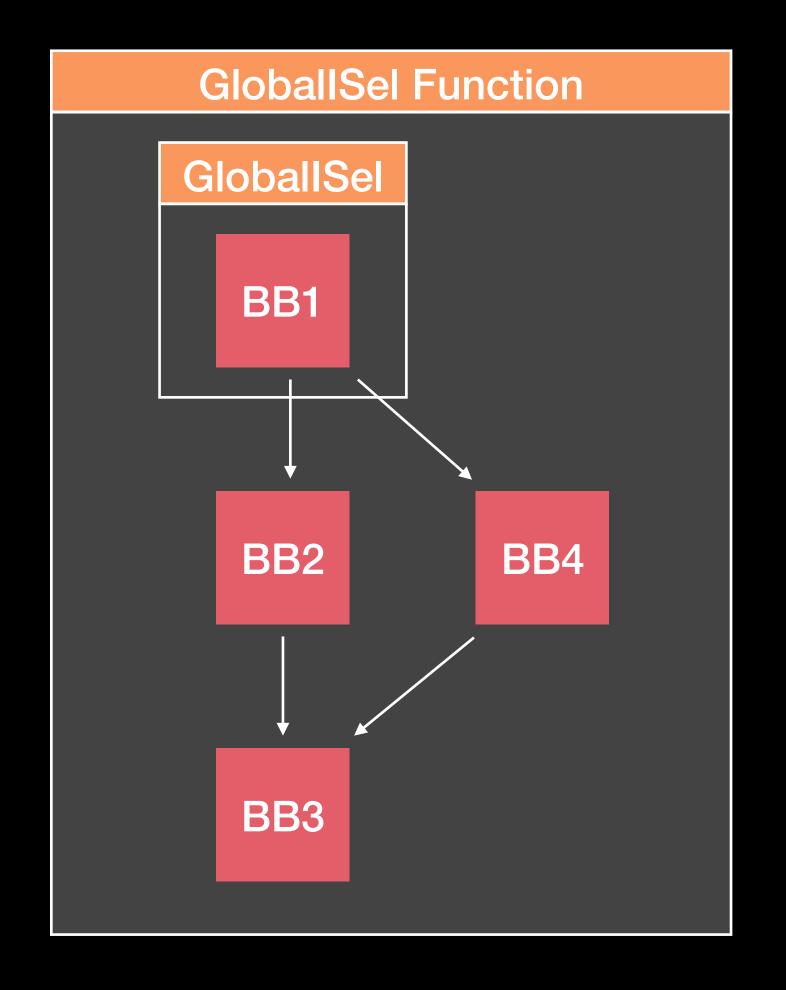
# Debugging

- It is error prone to implement optimizations from scratch
  - Special cases
  - Floating Point Precision issues (e.g. x \* y + z → fma(x, y, z))
  - Porting can be difficult too due to differences vs SelectionDAGISel
- It is especially hard to debug on GPUs
  - Xcode has tool to debug shaders, but it relies on the compiler being correct

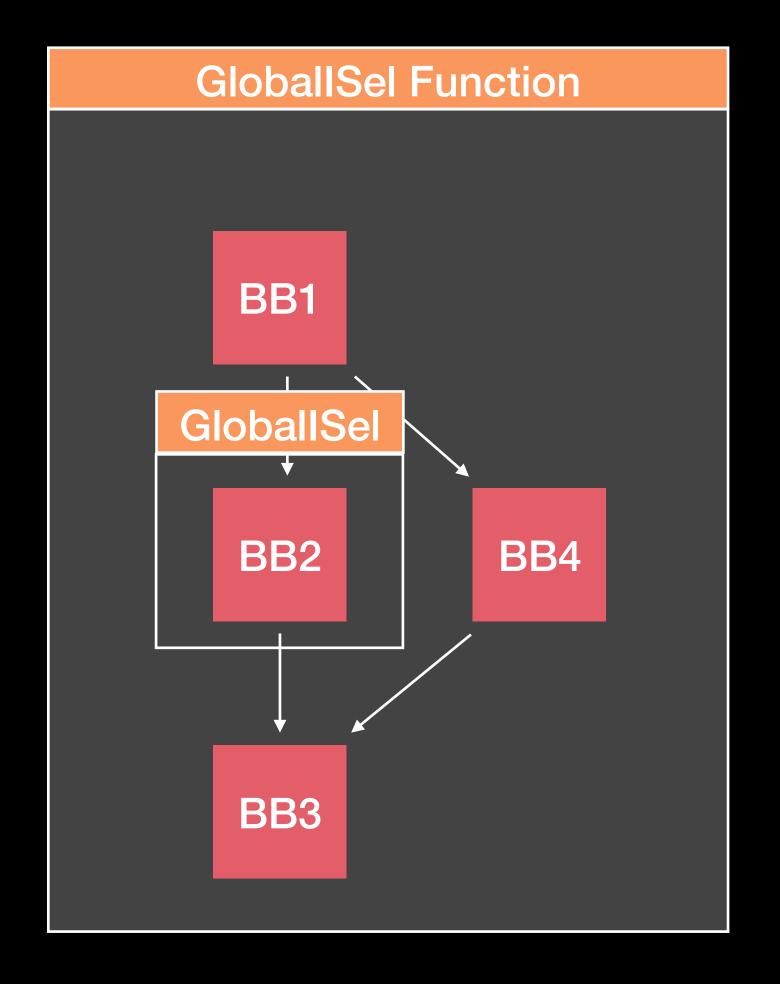
- LLVM Pass used by llvm-extract
- Promotes specified BasicBlocks to functions
- Exploitable to find critical block(s) for a bug
- GloballSel can be disabled per function



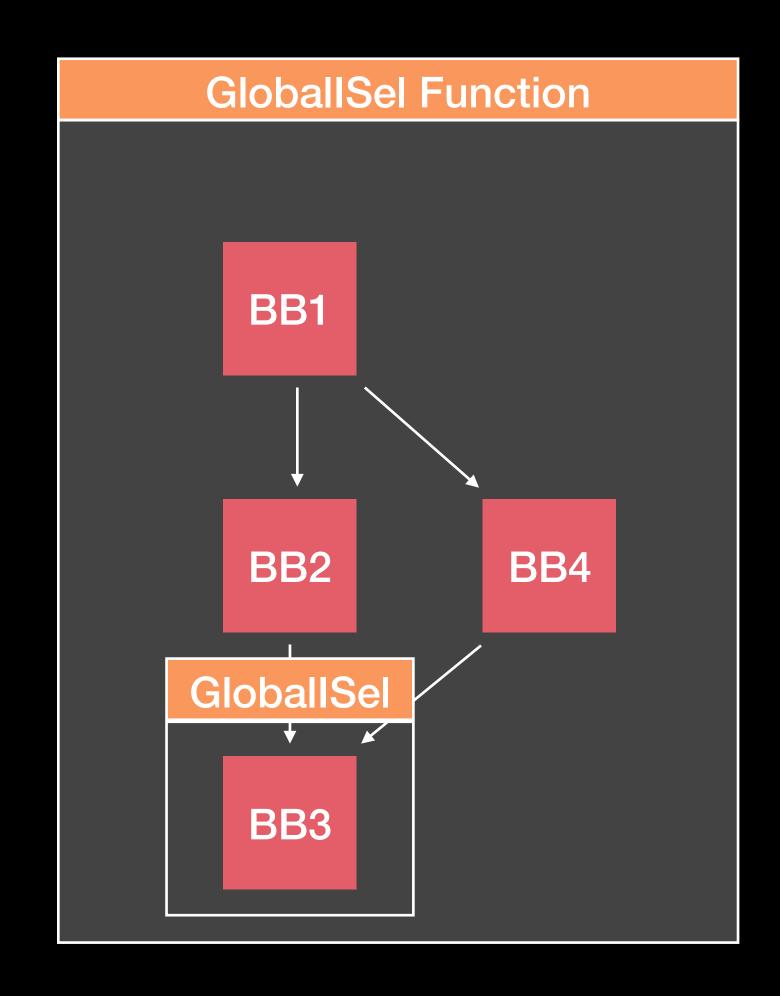
- LLVM Pass used by llvm-extract
- Promotes specified BasicBlocks to functions
- Exploitable to find critical block(s) for a bug
- GloballSel can be disabled per function



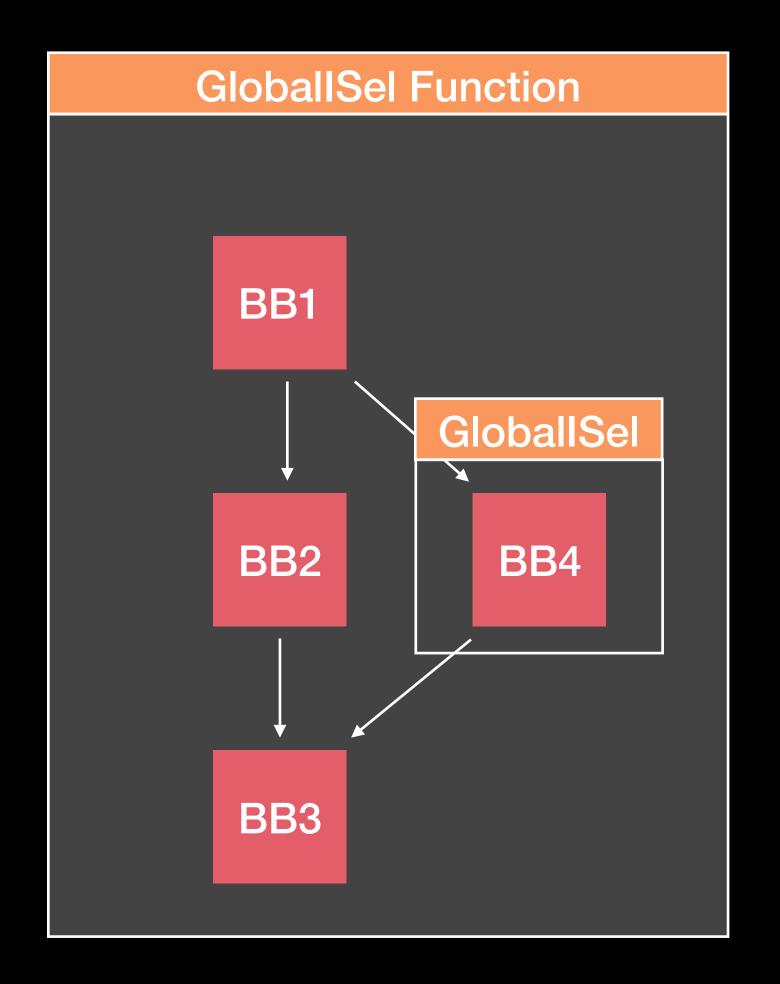
- LLVM Pass used by Ilvm-extract
- Promotes specified BasicBlocks to functions
- Exploitable to find critical block(s) for a bug
- GloballSel can be disabled per function



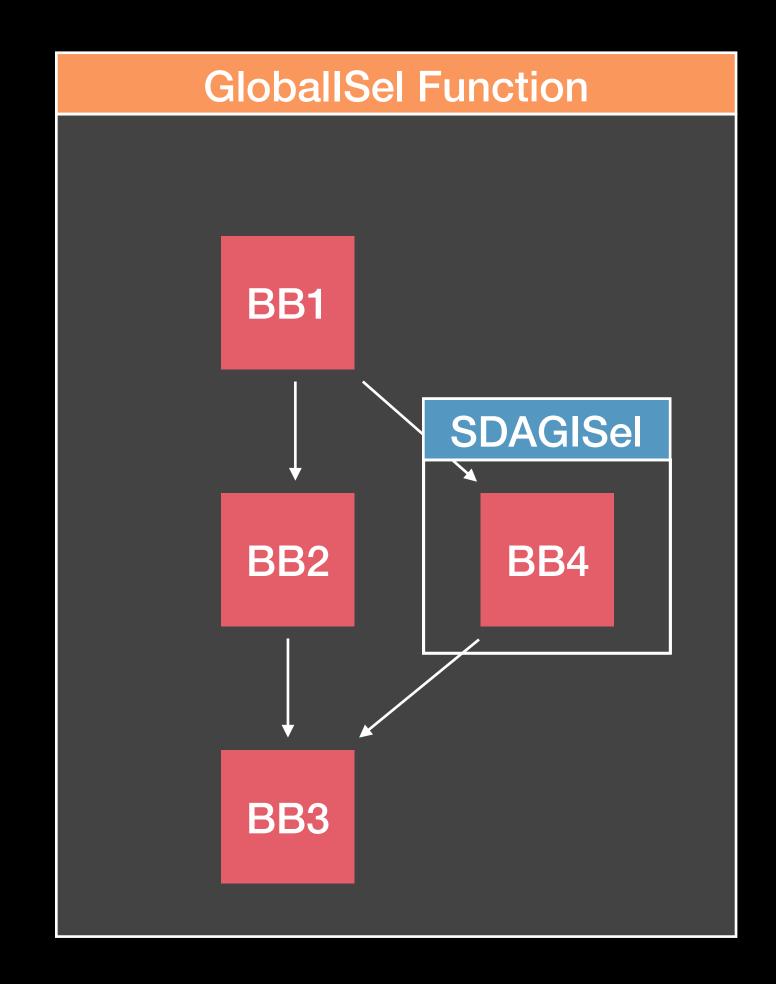
- LLVM Pass used by Ilvm-extract
- Promotes specified BasicBlocks to functions
- Exploitable to find critical block(s) for a bug
- GloballSel can be disabled per function



- LLVM Pass used by llvm-extract
- Promotes specified BasicBlocks to functions
- Exploitable to find critical block(s) for a bug
- GloballSel can be disabled per function

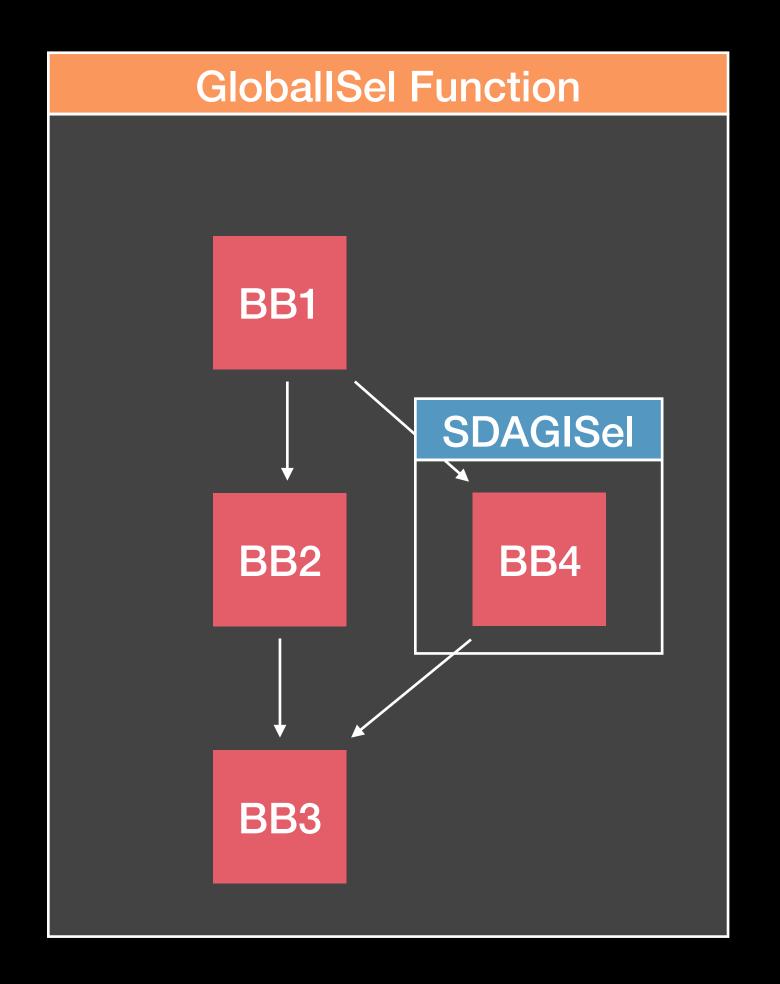


- LLVM Pass used by Ilvm-extract
- Promotes specified BasicBlocks to functions
- Exploitable to find critical block(s) for a bug
- GloballSel can be disabled per function



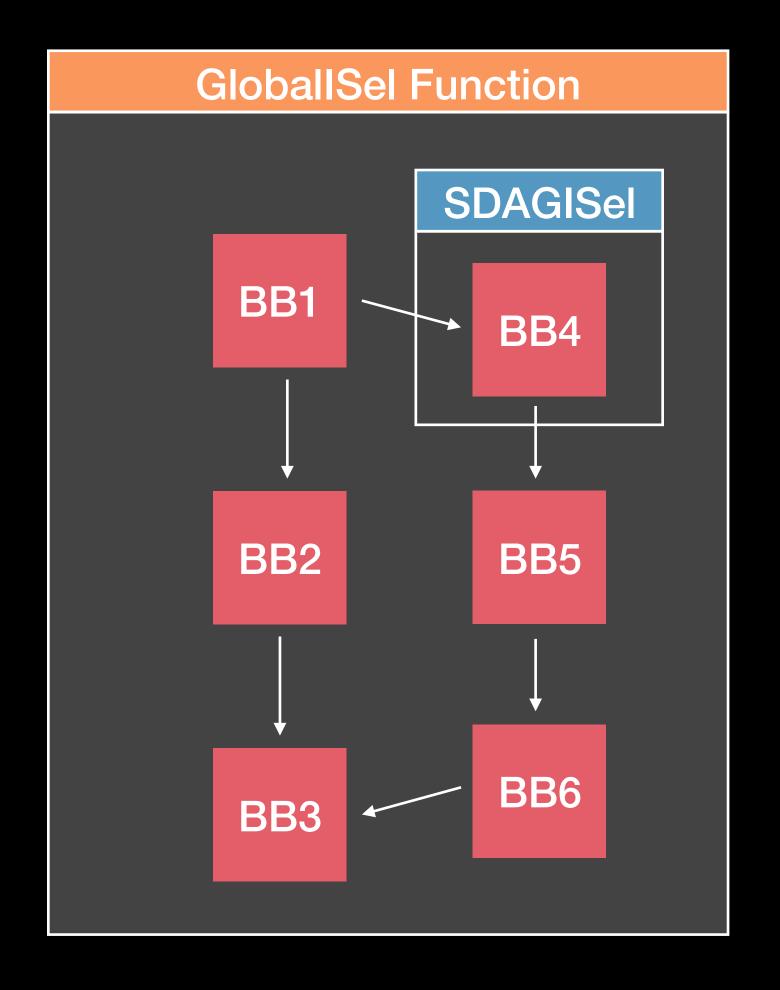


- Search space still too large?
  - Split the BasicBlocks and repeat



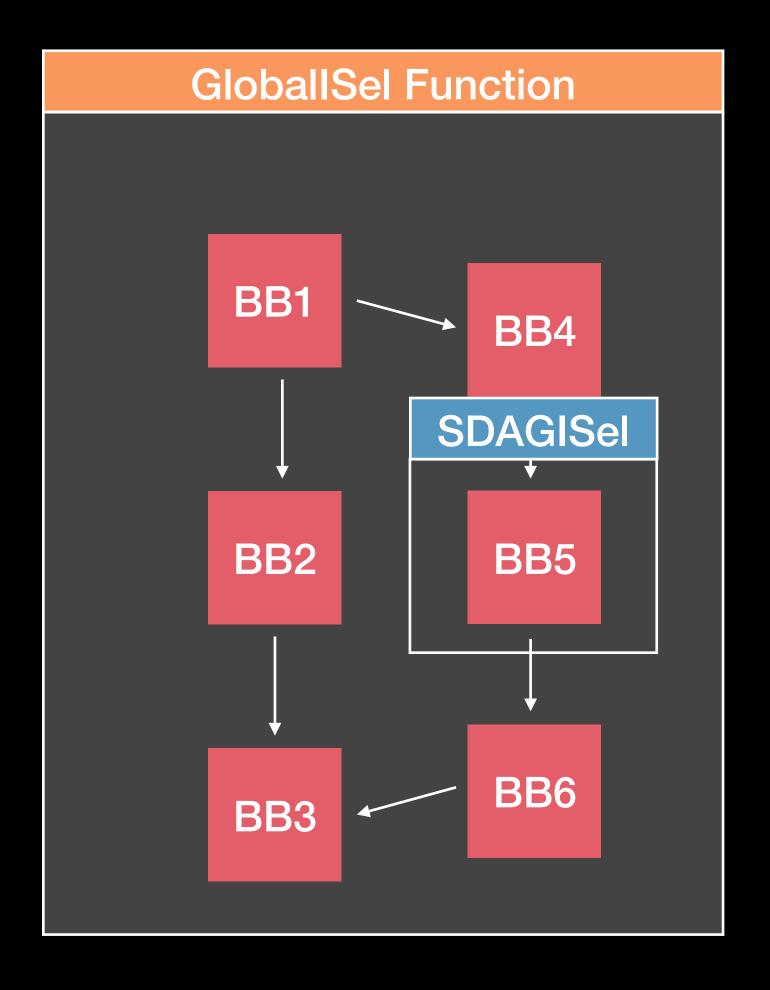


- Search space still too large?
  - Split the BasicBlocks and repeat



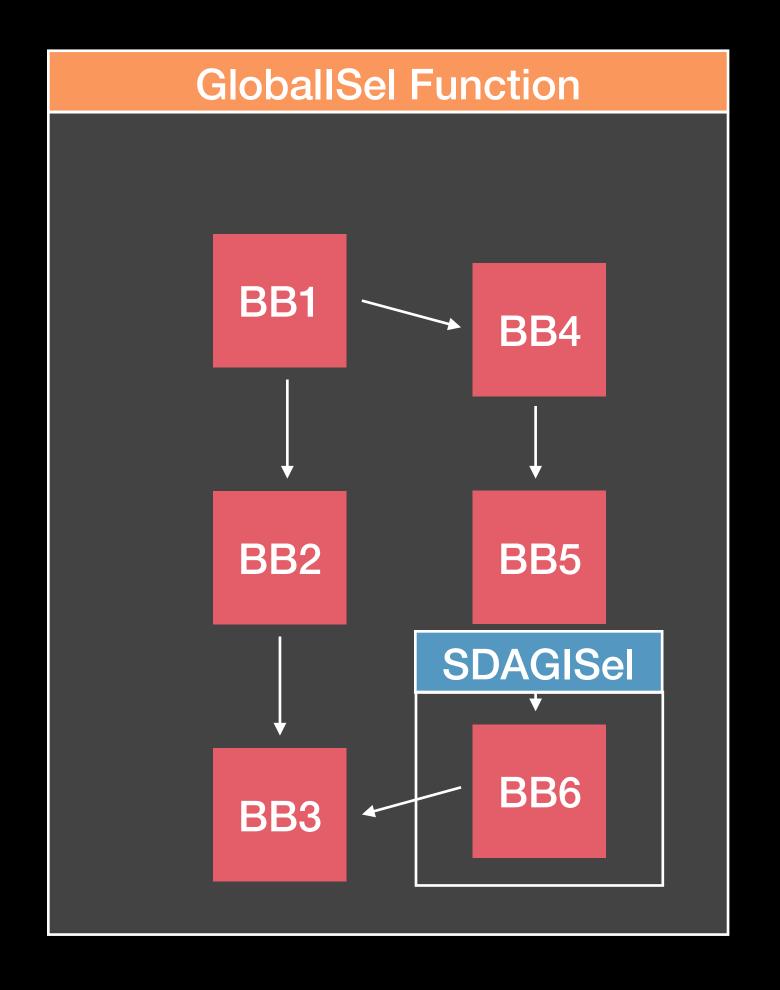


- Search space still too large?
  - Split the BasicBlocks and repeat



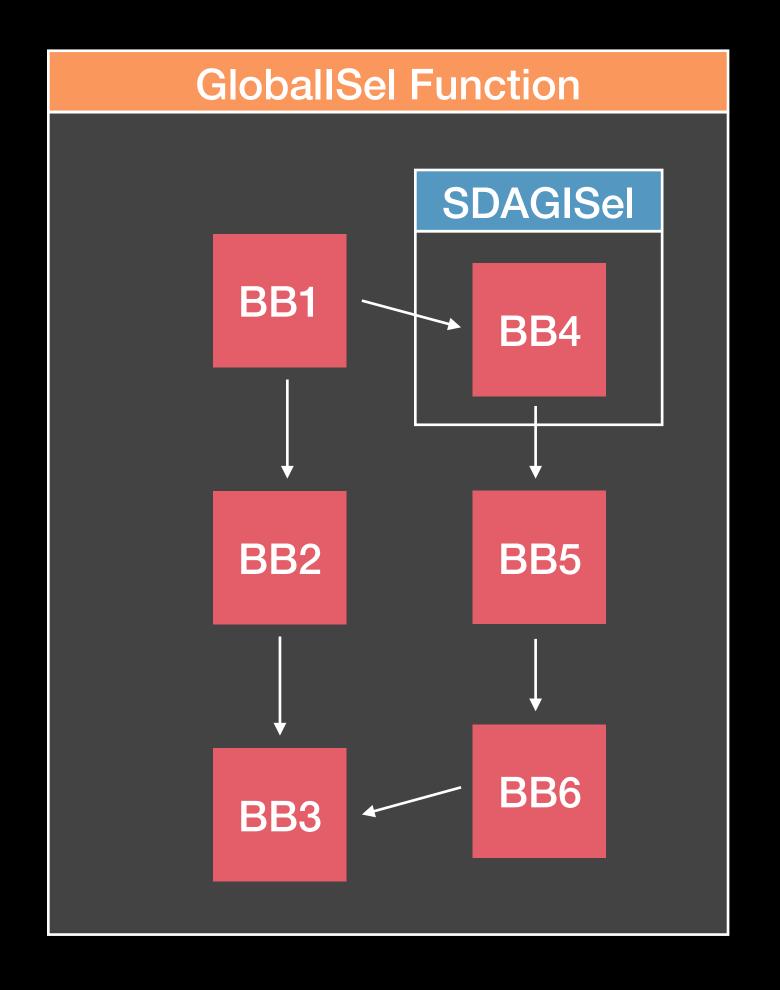


- Search space still too large?
  - Split the BasicBlocks and repeat





- Search space still too large?
  - Split the BasicBlocks and repeat





- All the components are upstream
- You will need a driver script to put them together

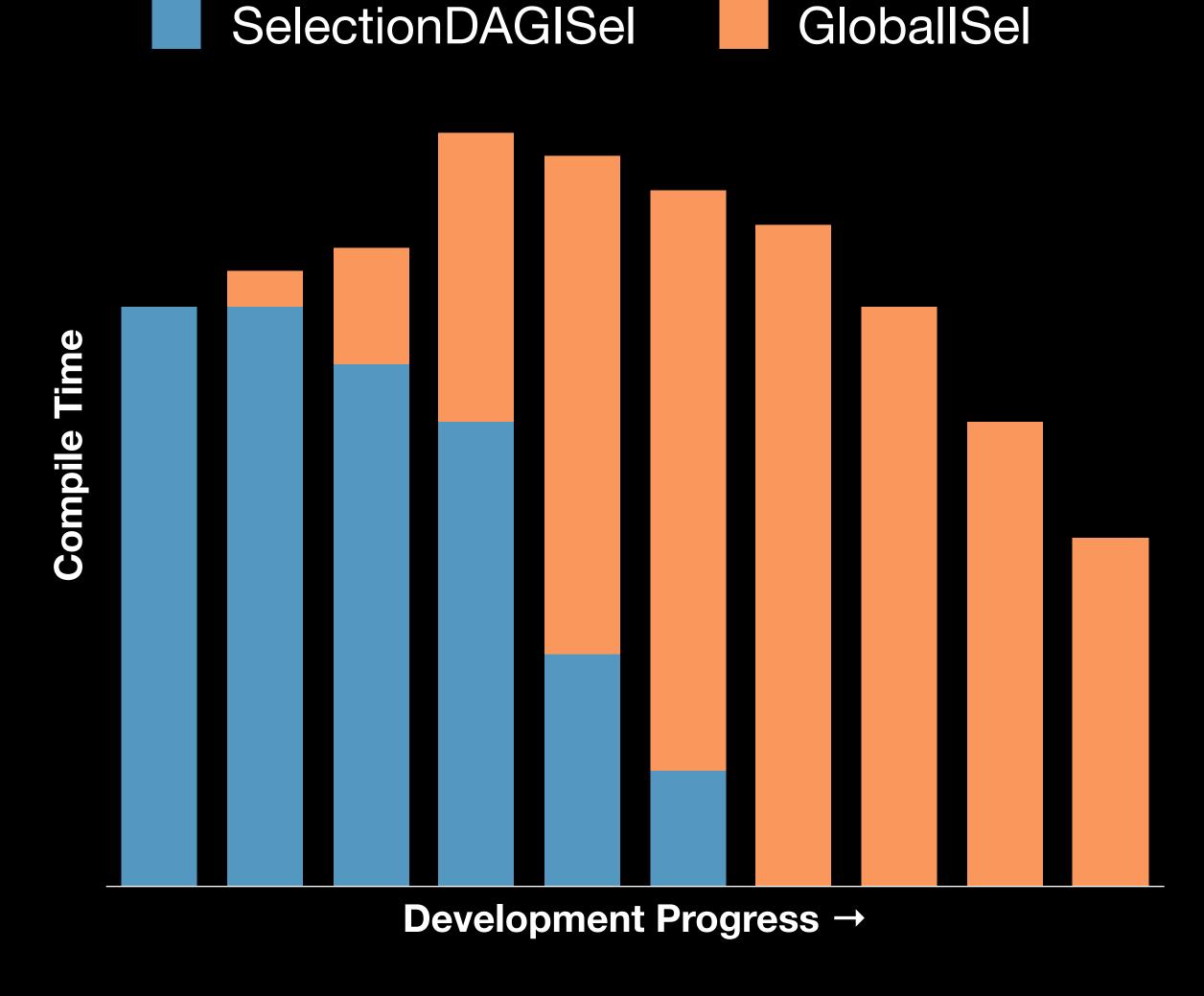
```
$ ./bin/llvm-extract -o - -S \
    -b 'foo:bb9;bb20' <input> > extracted.ll
```



# Advice

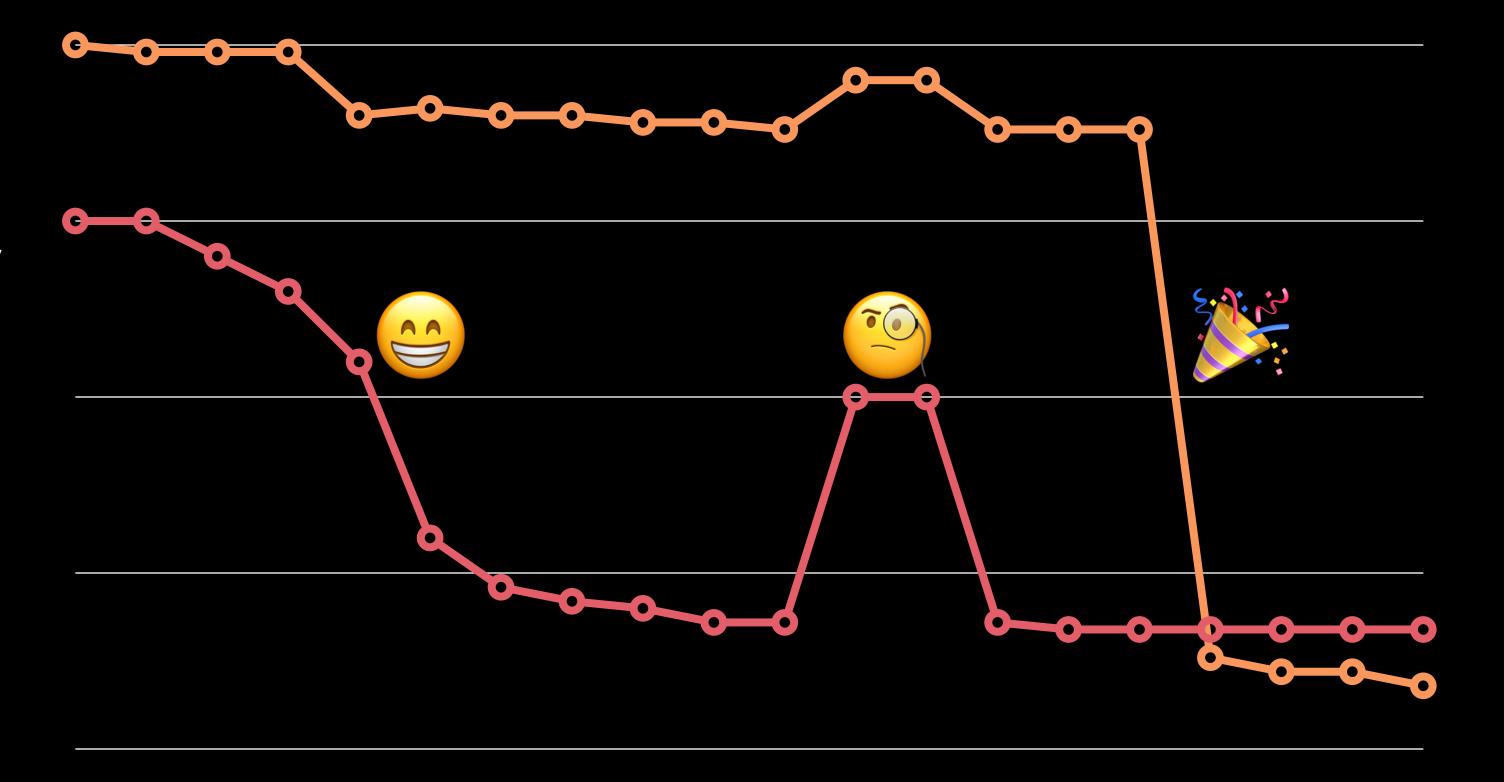
### Advice: Minimize Fallbacks

- Falling back:
  - Wastes compile time
  - Skews quality metrics



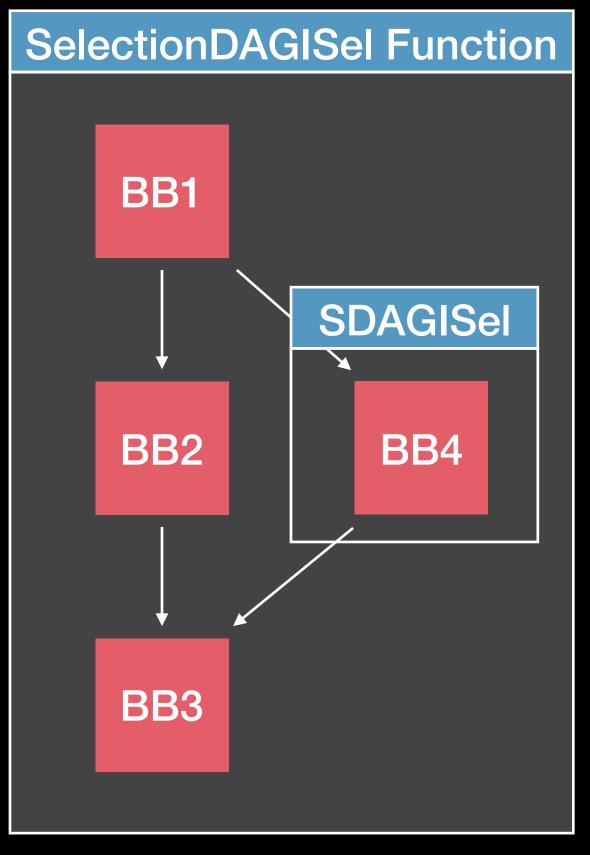
## Advice: Track Metrics Closely

- Catch regressions early
- Celebrate wins

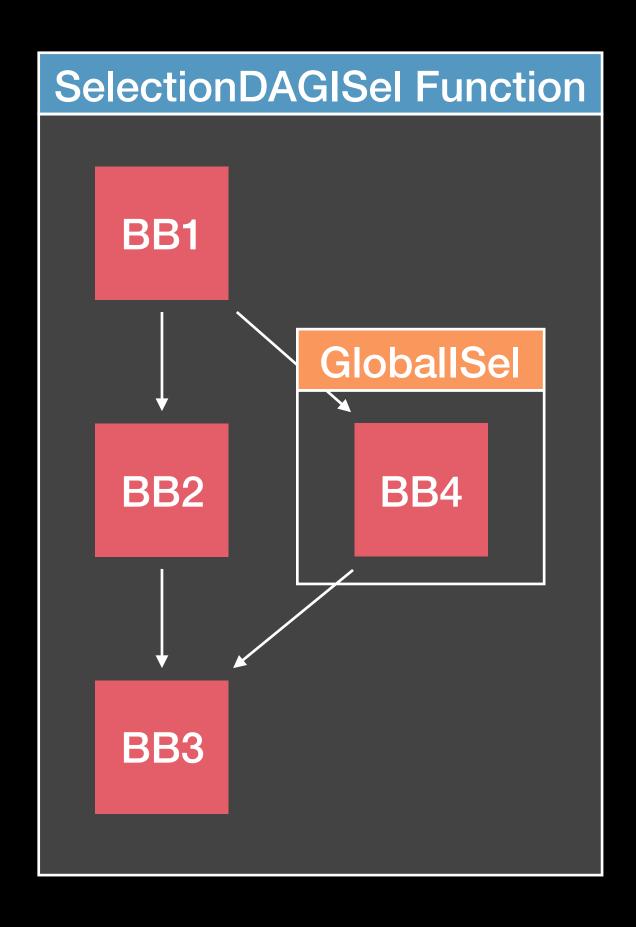


## Advice: Identify Key Optimizations

- Identify important optimizations
- Code Coverage Insights
- Minimize with BlockExtractor



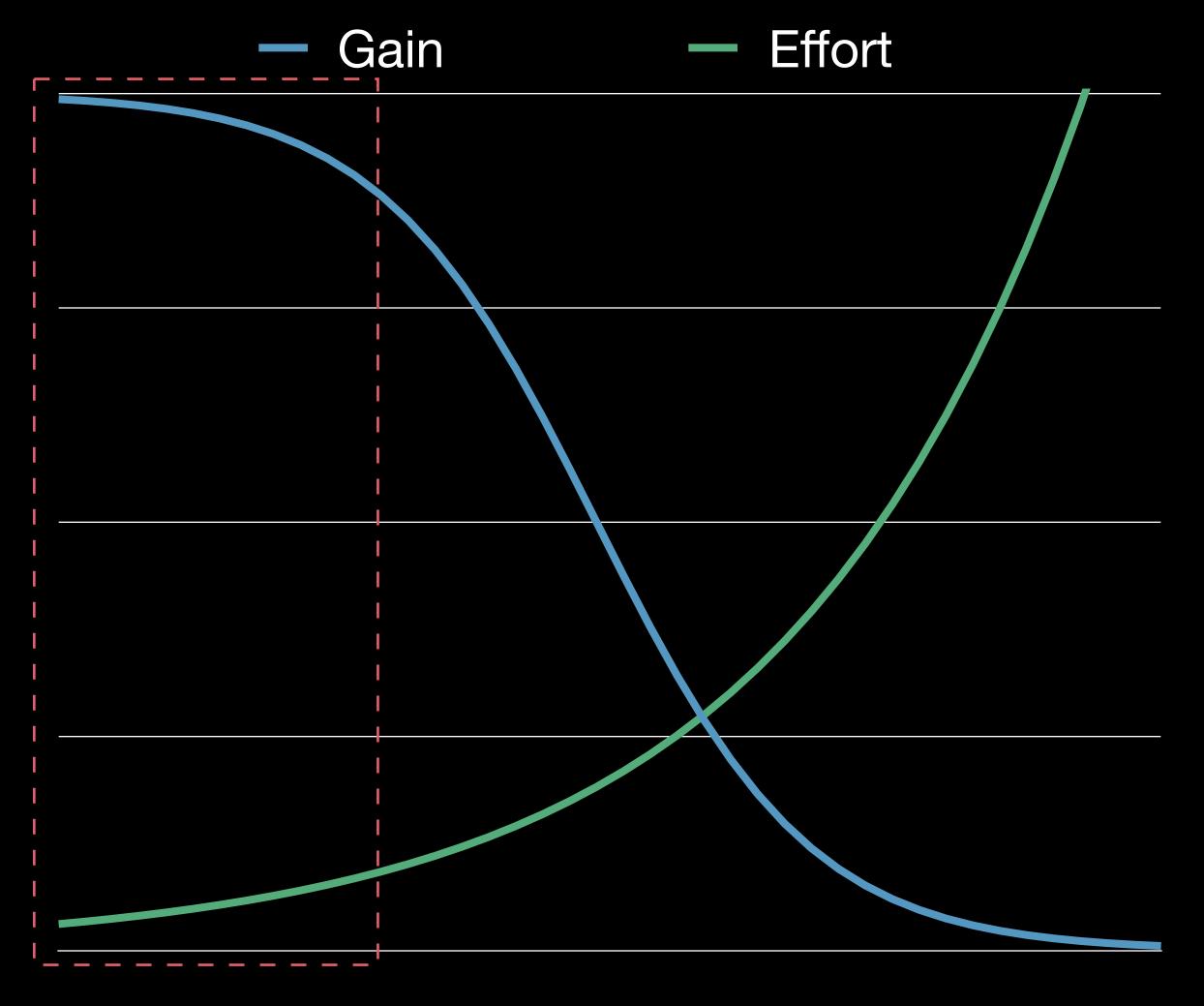




45 instrs, 5 due to BB4

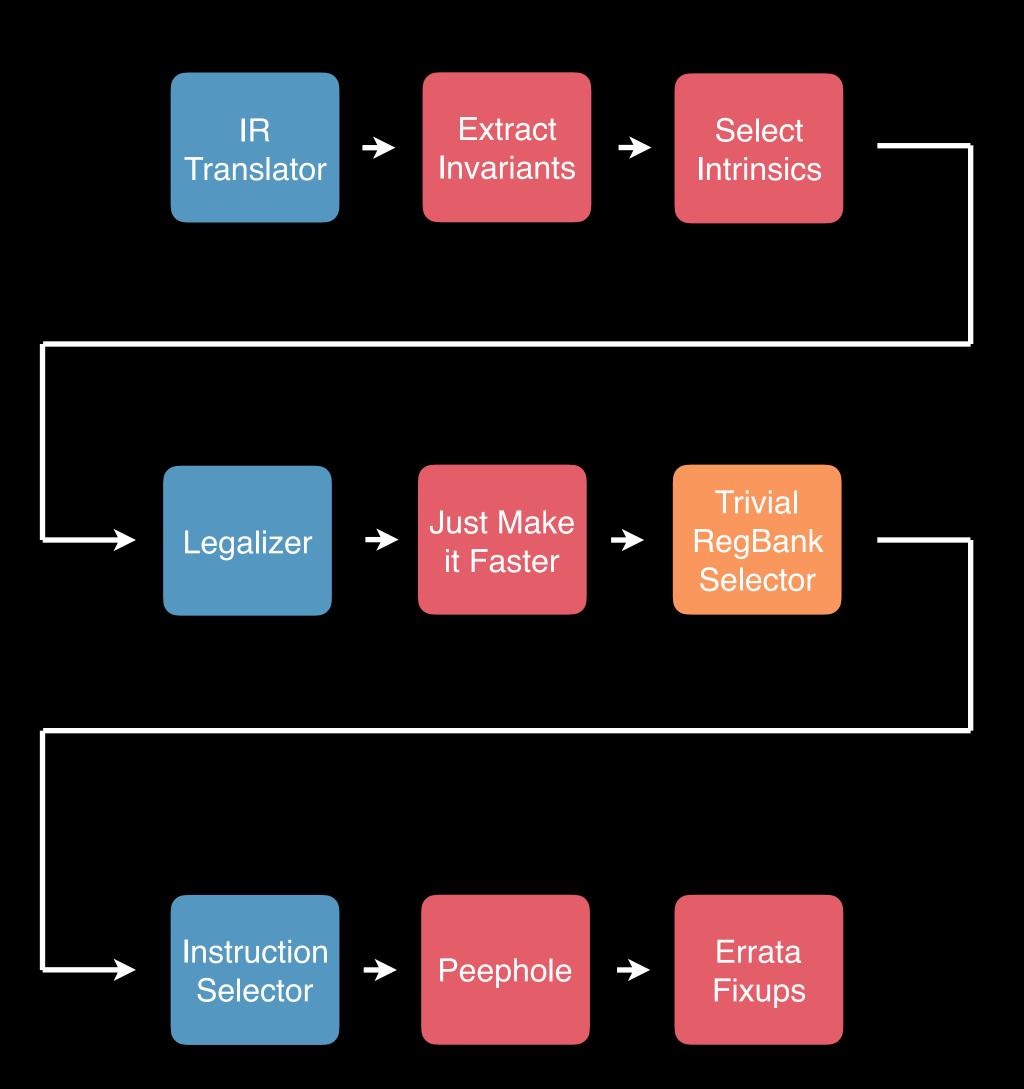
# Advice: Starting a Combiner

- Simple combines go a long way
- PreLegalizerCombiner and PostLegalizerCombiner are easy starting points



## Advice: Freedom

- Remember: Not a fixed pipeline
- Can replace passes
- Insert a pass where appropriate





# Work In Progress

### Declarative Combiner

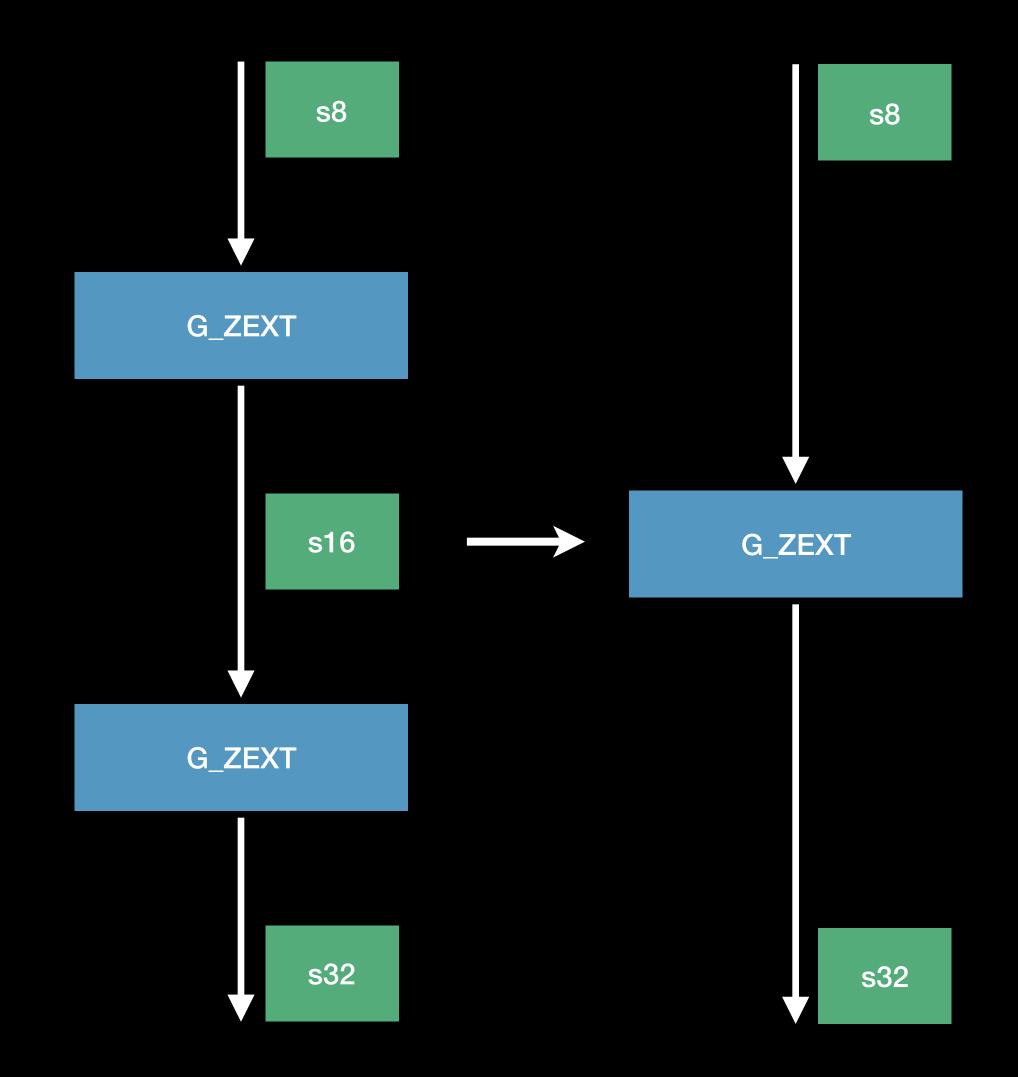
- Modify RuleSets
  - Targets may wish to disable rules or make them only apply in certain circumstances
- Analyze RuleSets
  - Enables various kinds of tooling
- Optimize RuleSets

### Goals

- Test Combine rules in isolation
- More debuggable: infinite loops and large rule-sets
- More target control
- Enable tools: profilers, coverage, static-analysis, proof engines
- Be independent of algorithm used

#### Declarative Rule

```
def : GICombineRule<
  (defs reg:$D, reg:$S),
  (match
      (G_ZEXT s16:$t1, s8:$S),
      (G_ZEXT s32:$D, s16:$t1)),
  (apply
      (G_ZEXT s32:$D, s8:$S))>;
```

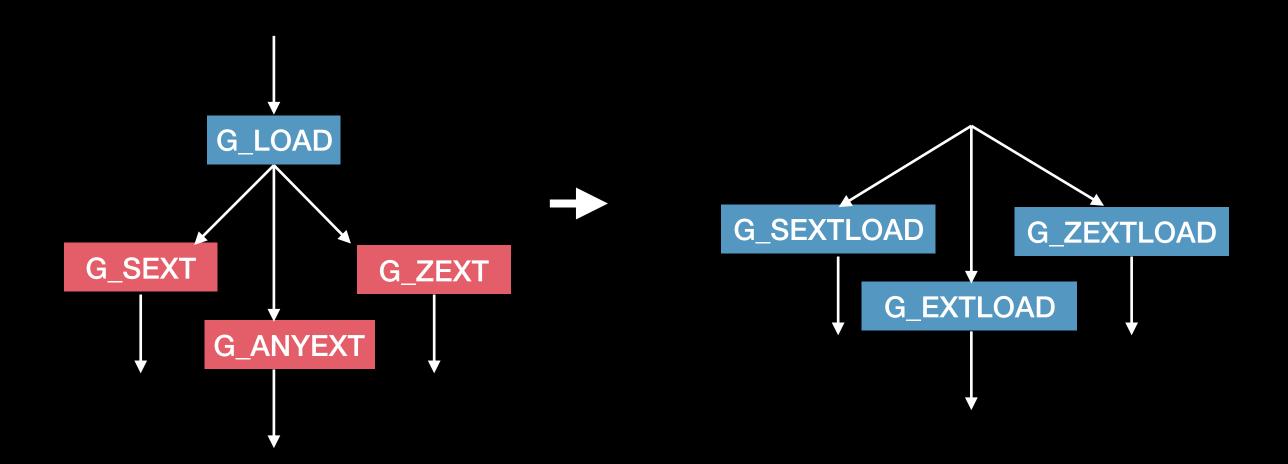


## Why not SelectionDAG's patterns?

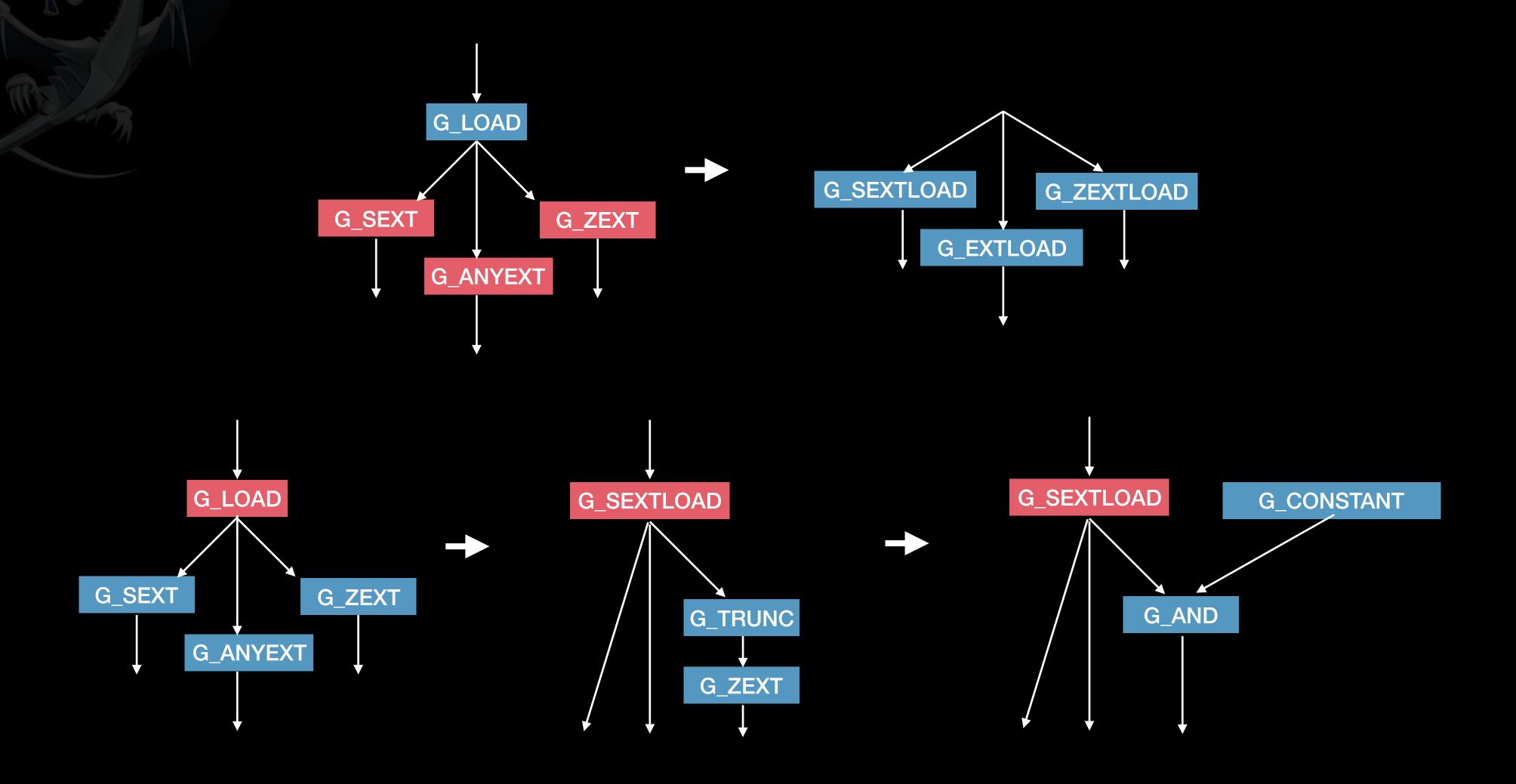
- They can't describe several classes of DAG.
  - Only the bottom-up tree-like DAG's with limited node sharing
- Can't describe multiple results from one instruction

## Example - SelectionDAG Style

- Analyzes def of the G\_SEXT, G\_ANYEXT, G\_ZEXT in isolation
- Folds the G\_LOAD down into the extend, duplicating the load
- Volatile/Atomics rejected unless hasOneUse()



# Example - GloballSel Style





## Debug Info

#### Rule Selection

- CombinerHelpers are declared in TableGen
- Specifies a class name and a list of combines in priority order
- Generated combiner ensures this order is honoured but still optimizes

```
def MyPreLegalizerCombinerHelper : GICombinerHelper<
   "MyGenPreLegalizerCombinerHelper",
   [copy_prop, fold_add_0, fold_mul_1, postpone_sext_for_add,
     postpone_zext_for_add, postpone_sext_for_sub, postpone_zext_for_sub,
     extending_loads]>;
```

#### Rule Selection

- CombineGroups may be specified to factor out:
  - Common combines (e.g. identities)
  - Common target features (e.g. unfused\_muladd, load\_multiple, bswap)

#### Rule Selection

- CombineGroups may be specified to factor out:
  - Common combines (e.g. identities)
  - Common target features (e.g. unfused\_muladd, load\_multiple, bswap)

```
def MyPreLegalizerCombinerHelper: GICombinerHelper<
    "MyGenPreLegalizerCombinerHelper",
    [trivial_combines, postpone_extends, extending_loads]>;
```

### Rule Selection

- Generated combiner includes a command line option when asked
  - -myprelegalizercombiner-disable-rule=1
  - ► -myprelegalizercombiner-disable-rule=0-50,75-100
  - -myprelegalizercombiner-disable-rule=fold\_2\_plus\_2\_to\_5

```
def MyPreLegalizerCombinerHelper: GICombinerHelper<
   "MyGenPreLegalizerCombinerHelper",
   [trivial_combines, postpone_extends, extending_loads]> {
   let DisableRuleOption = "myprelegalizercombiner-disable-rule";
}
```

### Rule Selection

- Sometimes we can generally use a group but there's a small flaw
- Combiners (and maybe groups in future) can modify their contents
- Exact modifiers TBD

## Integration

- Generates a Combiner
- Integrate into CombinerInfo via constructor and combine() tweak

```
AArch64GenPreLegalizerCombinerHelper Generated;
if (!Generated.parseCommandLineOption())
    report_fatal_error("Invalid rule identifier");

if (Generated.tryCombineAll(Observer, MI, B))
    return true;
```

### Extensibility

## Development Tools

- Coverage Are rules tested? Do they trigger in practice?
- Profiler Are they worth their cost?
- Controlled application of rules?
  - If I applied them in this order would the outcome be better?

# Debugging Tools

- Rule Bisection Which one caused a miscompilation?
- N-stable Loop Detection Why doesn't my combiner terminate?
- Rule Proving, i.e. ALIVE for backend? Are my rules correct?
- State machine debugger? What is the combiner doing?
- MIR Patches Re-construct intermediate MIR





In 2017, we got GloballSel working but code quality wasn't there yet

- In 2017, we got GloballSel working but code quality wasn't there yet
- Added continuous CSE, Combine, KnownBits, and SimplifyDemandedBits

- In 2017, we got GloballSel working but code quality wasn't there yet
- Added continuous CSE, Combine, KnownBits, and SimplifyDemandedBits
- By 2019:
  - Compile time 45% faster than SelectionDAGISel
  - Generated code quality on par with SelectionDAGISel

- In 2017, we got GloballSel working but code quality wasn't there yet
- Added continuous CSE, Combine, KnownBits, and SimplifyDemandedBits
- By 2019:
  - Compile time 45% faster than SelectionDAGISel
  - Generated code quality on par with SelectionDAGISel
- Other targets are actively working on GloballSel

- In 2017, we got GloballSel working but code quality wasn't there yet
- Added continuous CSE, Combine, KnownBits, and SimplifyDemandedBits
- By 2019:
  - Compile time 45% faster than SelectionDAGISel
  - Generated code quality on par with SelectionDAGISel
- Other targets are actively working on GloballSel
- We shipped it! You might even be using it!

