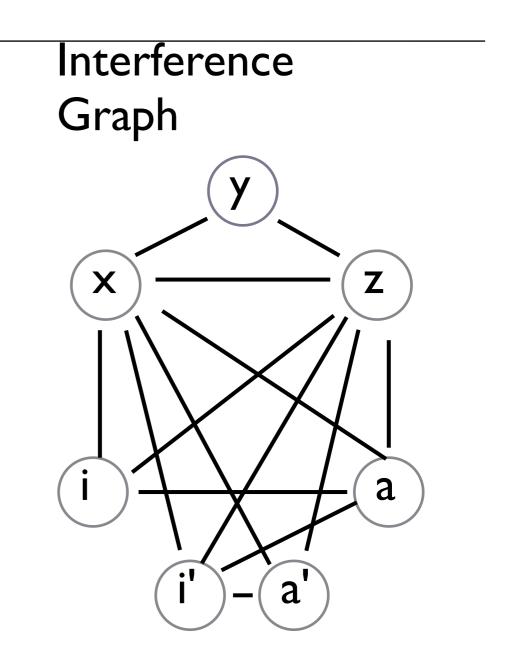
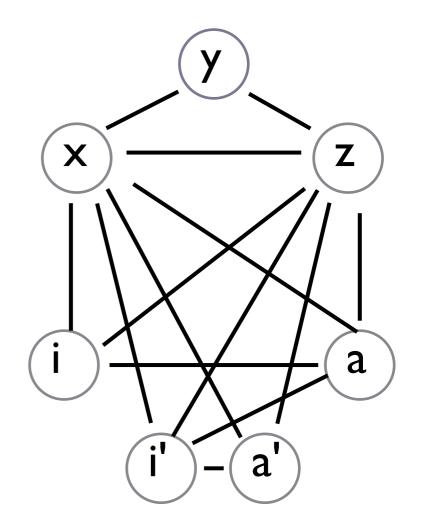
Graph Coloring Register Allocation

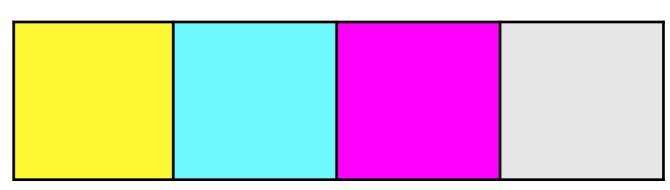
Register Allocation

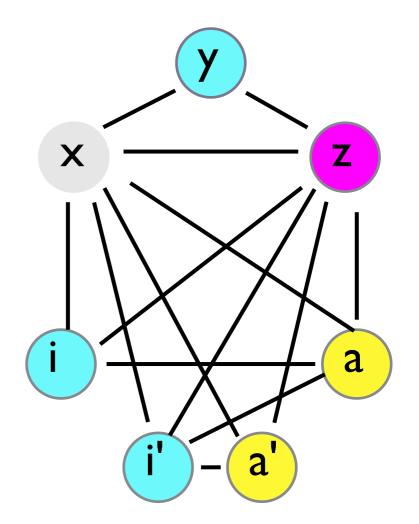
3(.5) Steps

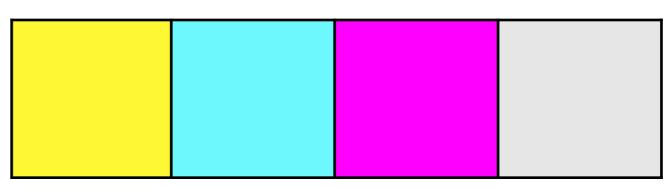
- Liveness analysis: identify when each variable's value is needed in the program
- 2. **Conflict analysis**: identify which variables interfere with each other
- 3. **Graph Coloring**: assign variables to registers so that interfering registers are assigned different registers.
 - Spilling: if necessary, assign some variables to stack slots

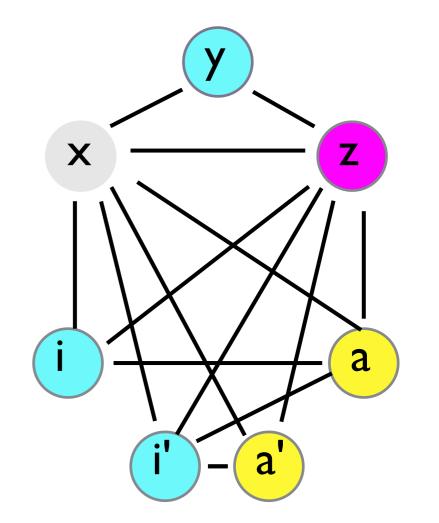


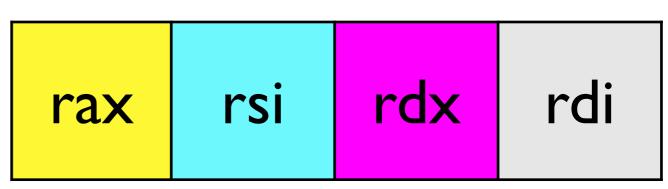




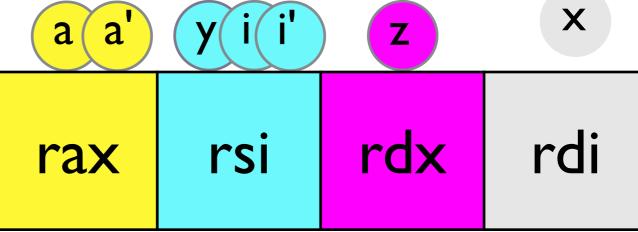








```
def f(x,y,z):
  def loop(i,a):
    if i == 0:
      a * z
    else:
      let i' = i - 1 in
      let a' = a + x in
      icall(loop; i', a')
  end
                                         Ζ
                              a
                            a
  icall(loop; y, 0)
```



	f:			
Example	m	ov rax	K, 0	
<pre>def f(x,y,z): def loop(i,a): if i == 0: a * z else: let i' = i - 1 in let a' = a + x in icall(loop; i', a') end icall(loop; y, 0)</pre>	j i r els s	mp rsing for the set of the set o	5 ax, rdx i, 1 k, rdi op z	X
	rax	rsi	rdx	rdi

Graph Coloring Register Allocation

Given our register conflict graph, want to assign a register to each variable so that no interfering variables are assigned the same register.

Graph Coloring Register Allocation

Given our register conflict graph, want to assign a register to each variable so that no interfering variables are assigned the same register.

Equivalent to graph coloring of the interference graph

• think of each register as a "color" and we want to paint each node so that no adjacent nodes are the same color.

Graph Coloring Register Allocation

Given our register conflict graph, want to assign a register to each variable so that no interfering variables are assigned the same register.

Equivalent to graph coloring of the interference graph

• think of each register as a "color" and we want to paint each node so that no adjacent nodes are the same color.

Efficient algorithm for graph coloring -> efficient algorithm for graph coloring!

Graph Coloring is Hard

Determining a whether a graph is k-colorable is NPcomplete for k > 2.

• So no polytime algorithm is known

Graph Coloring is Hard

Determining a whether a graph is k-colorable is NPcomplete for k > 2.

• So no polytime algorithm is known

Does that mean register allocation is NP-hard?

Chaitin et al, "Register allocation via coloring", Computer Languages 1981

 Showed that the register allocation problem for a language with assignments and arbitrary control flow (goto) is equivalent to graph coloring

Chaitin et al, "Register allocation via coloring", Computer Languages 1981

- Showed that the register allocation problem for a language with assignments and arbitrary control flow (goto) is equivalent to graph coloring
- every graph arises as the interference graph of some program

So register allocation of their language is NP complete.

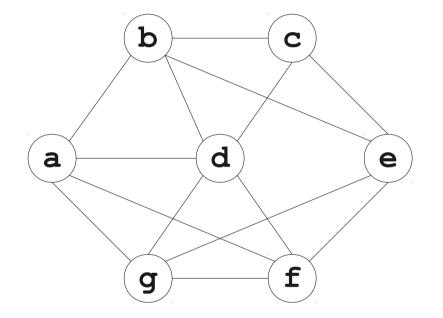
Chaitin et al, "Register allocation via coloring", Computer Languages 1981

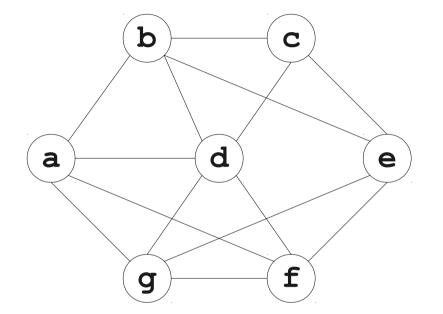
- Showed that the register allocation problem for a language with assignments and arbitrary control flow (goto) is equivalent to graph coloring
- every graph arises as the interference graph of some program

So register allocation of their language is NP complete.

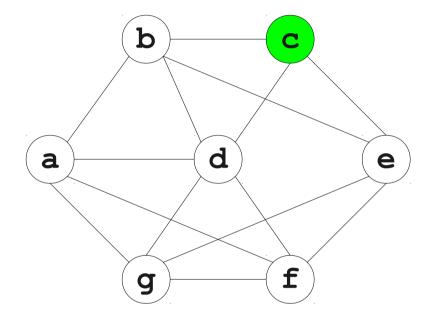
 But our programs are more restrictive: Functional/SSA form...we'll come back to this

- Intuition:
 - Suppose we are trying to *k*-color a graph and find a node with fewer than *k* edges.
 - If we delete this node from the graph and color what remains, we can find a color for this node if we add it back in.
 - Reason: With fewer than *k* neighbors, some color must be left over.
- Algorithm:
 - Find a node with fewer than *k* outgoing edges.
 - Remove it from the graph.
 - Recursively color the rest of the graph.
 - Add the node back in.
 - Assign it a valid color.

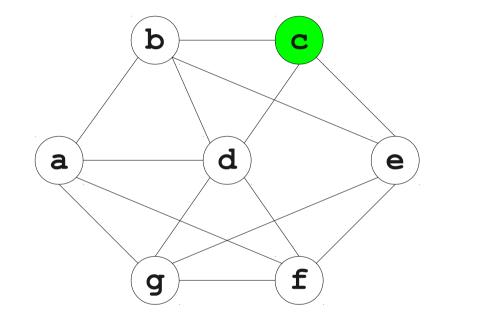






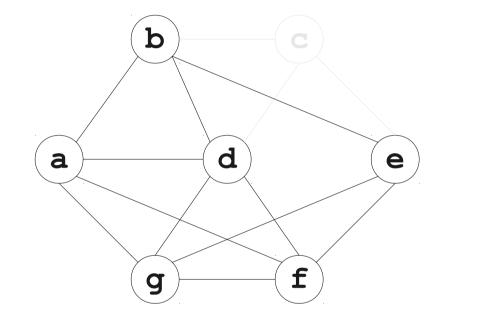






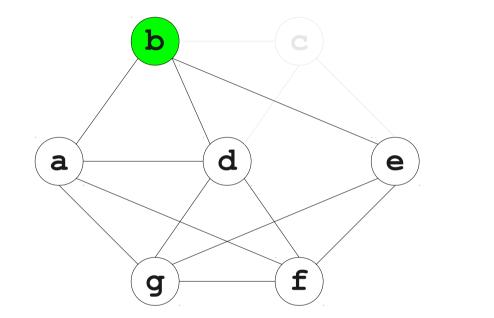


С



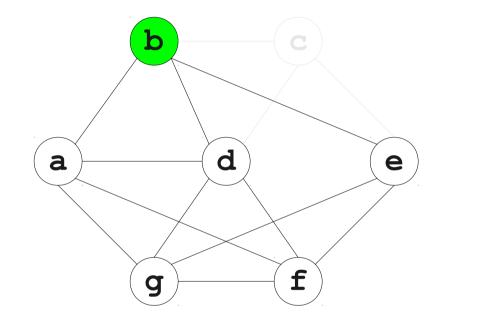


С





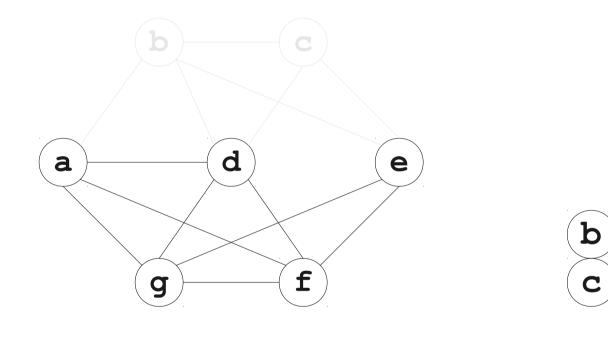




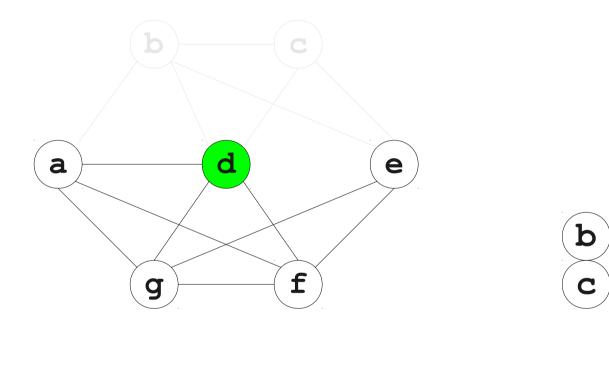


b

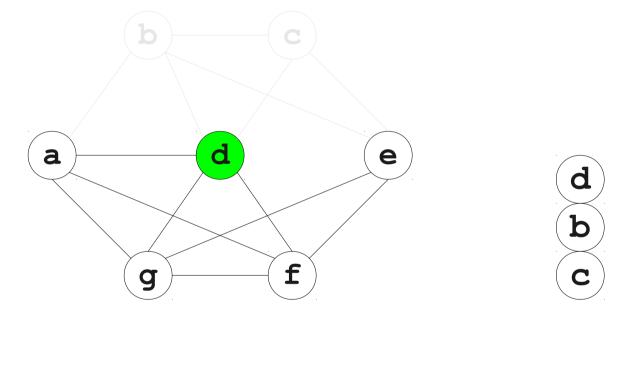
С



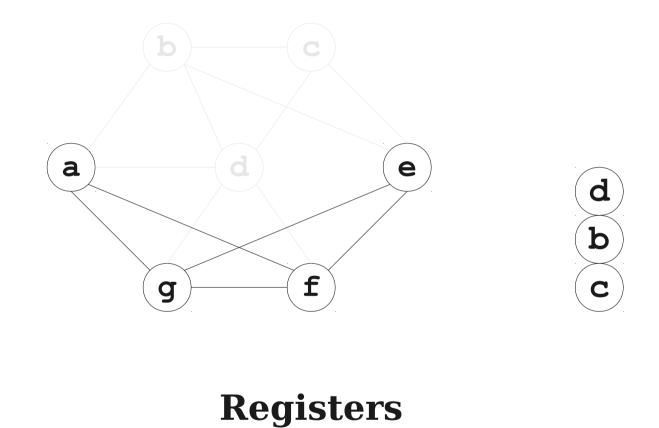










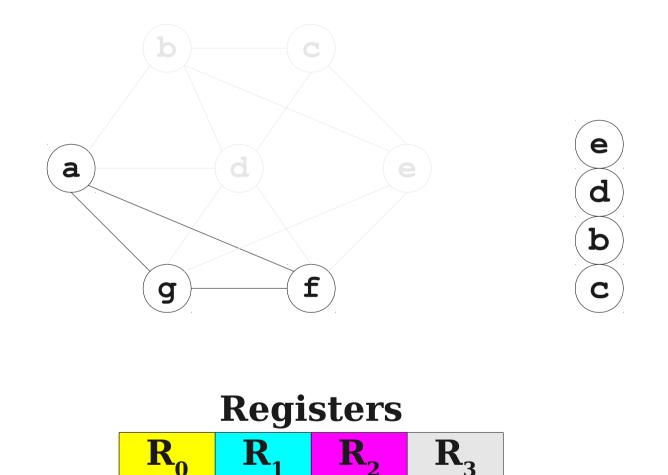


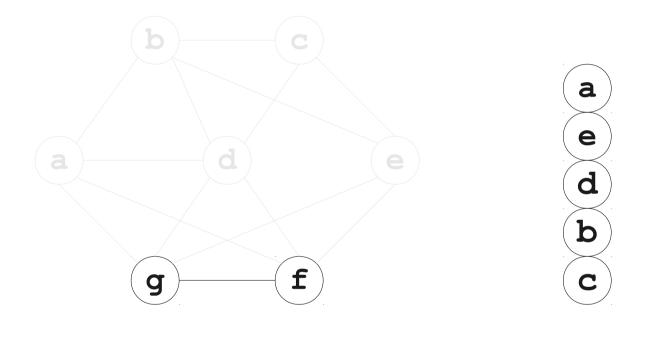
 \mathbf{R}_2

R₃

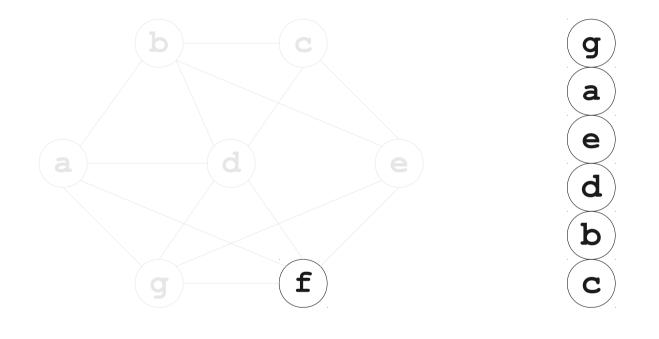
R₀

R₁

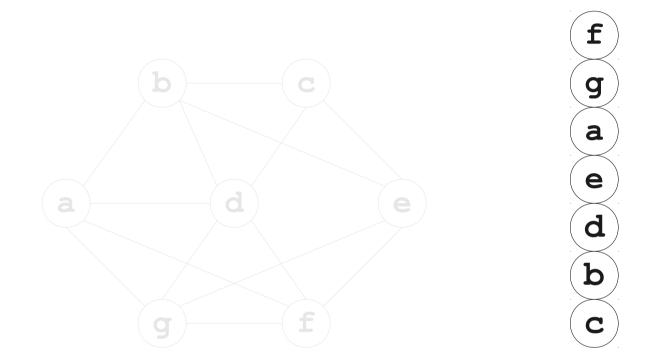




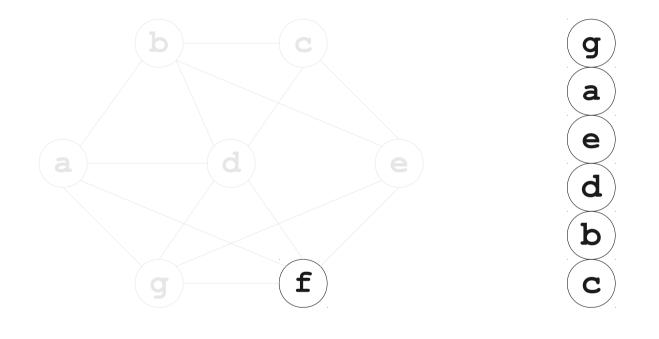




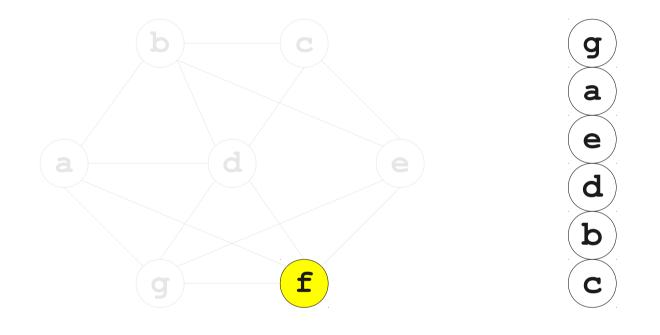




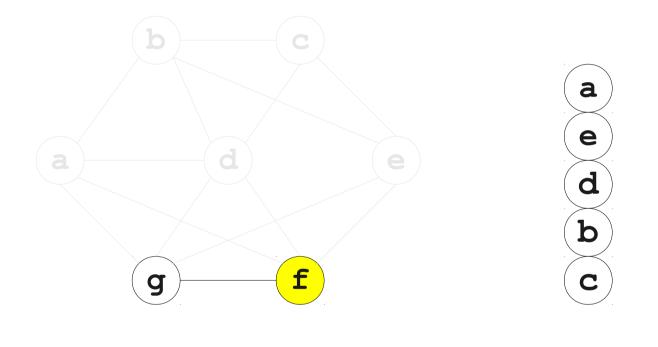




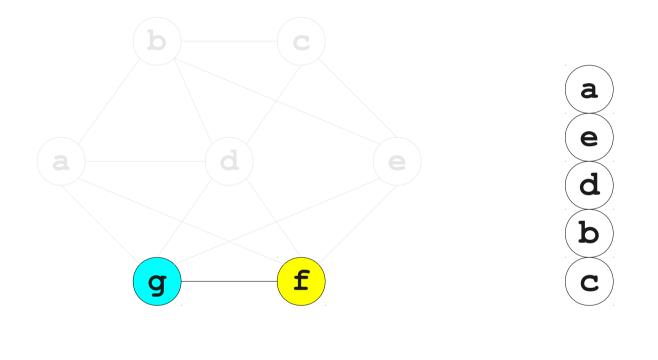




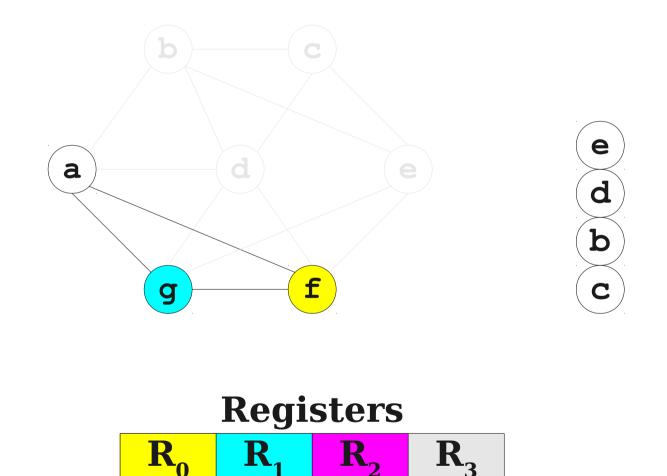


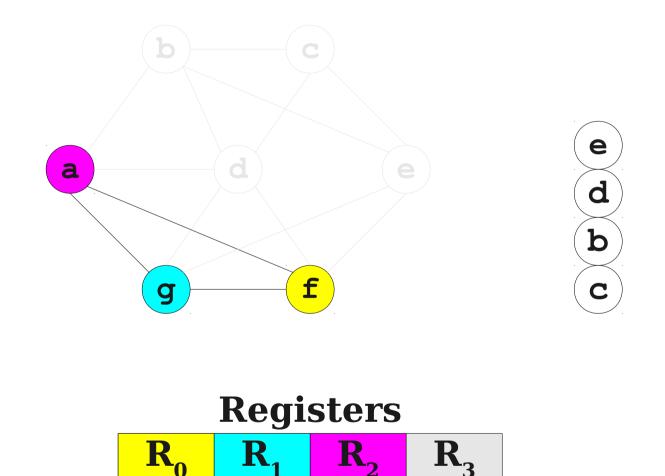


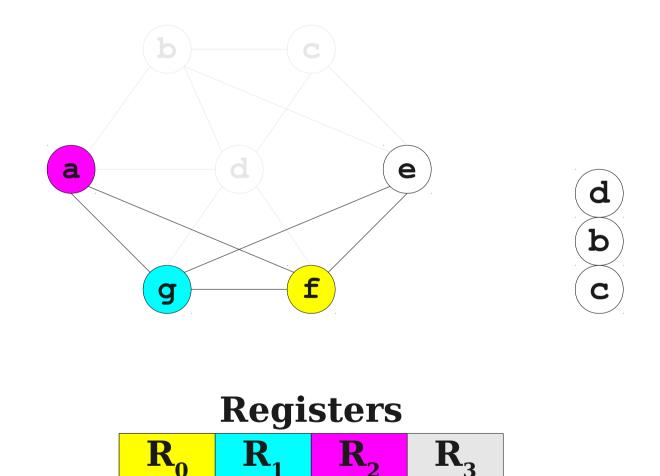


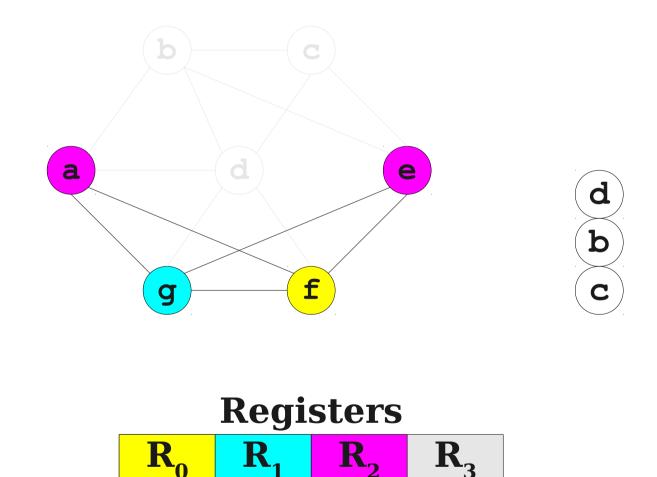


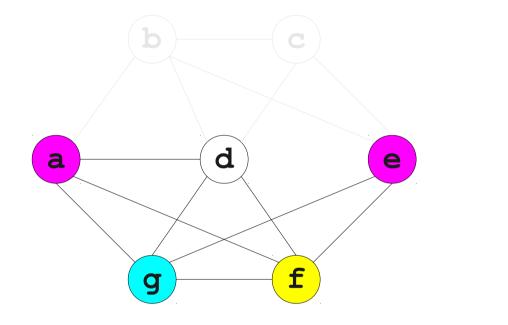








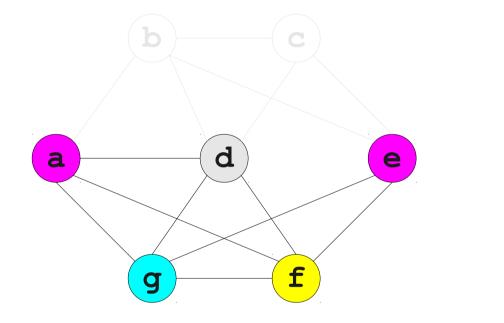






b

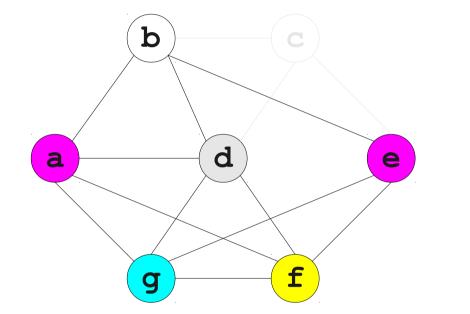
С





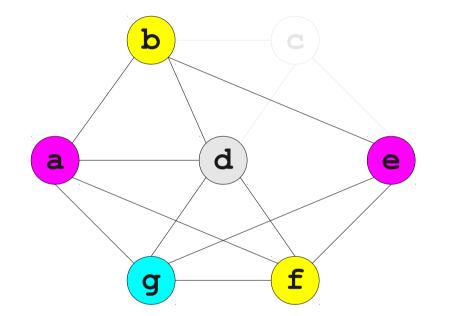
b

С



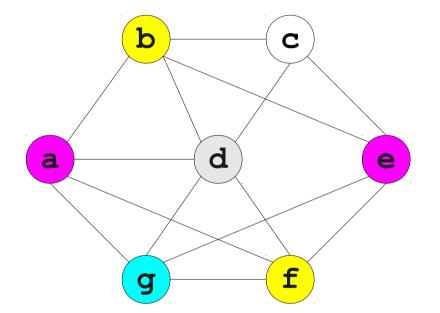




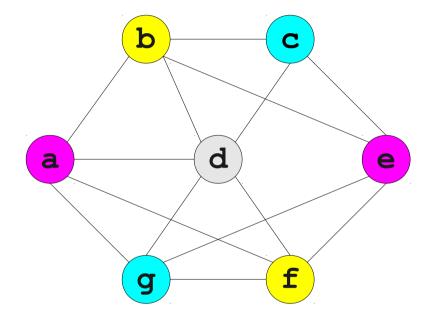








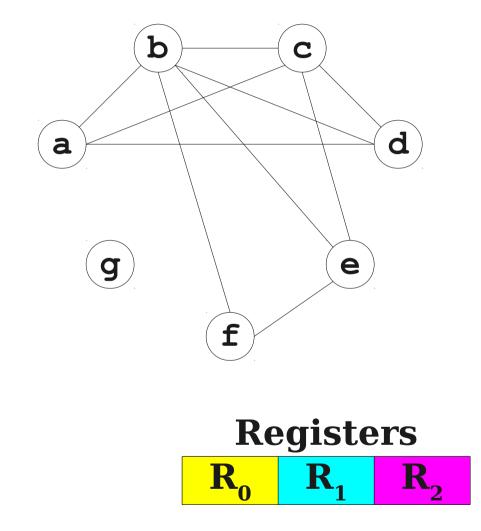


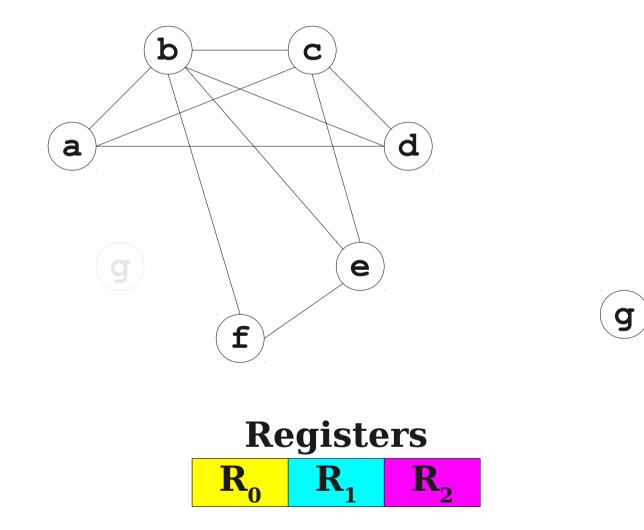


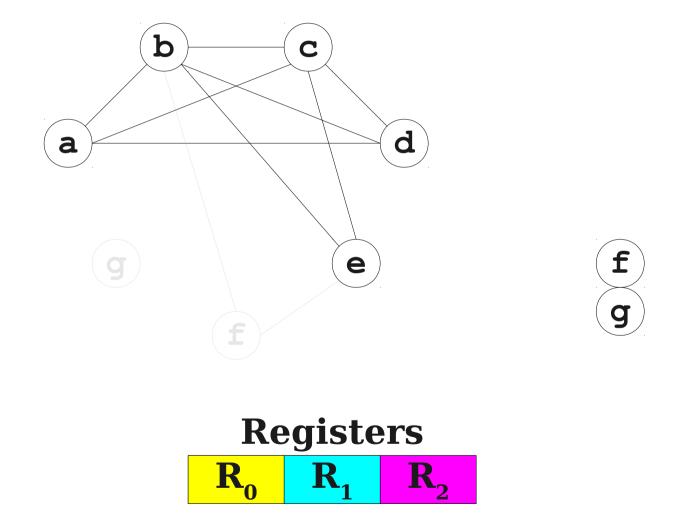


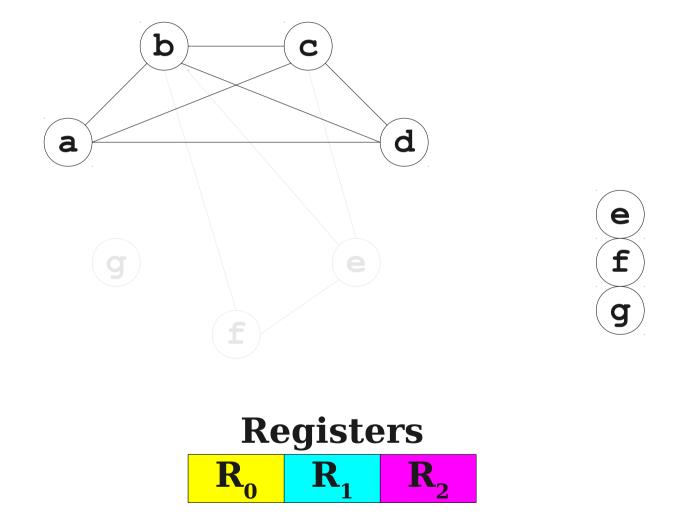
One Problem

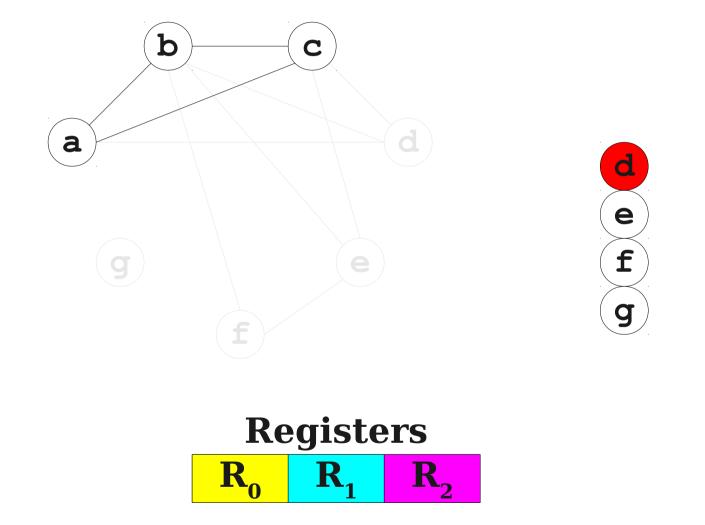
- What if we can't find a node with fewer than k neighbors?
- Choose and remove an arbitrary node, marking it "troublesome."
 - Use heuristics to choose which one.
- When adding node back in, it may be possible to find a valid color.
- Otherwise, we have to spill that node.

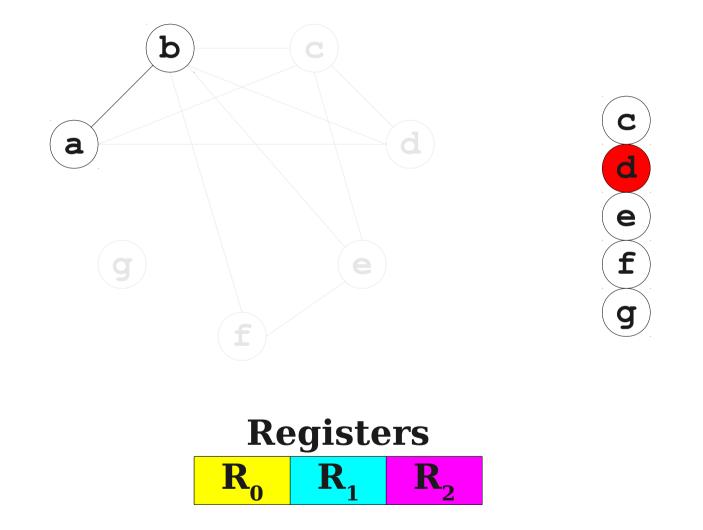


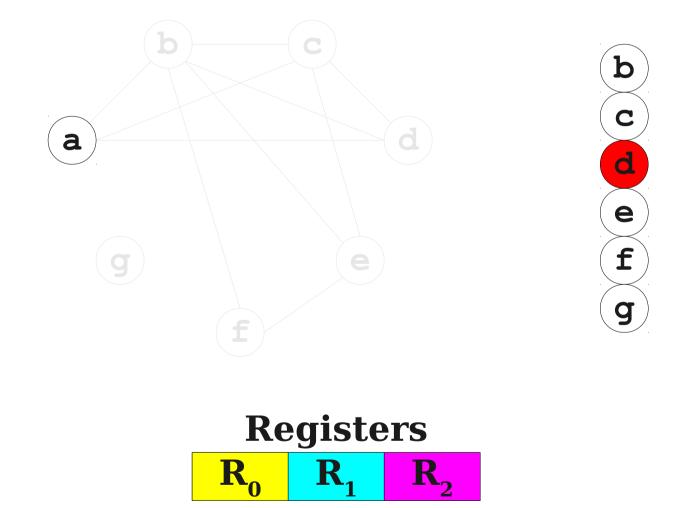






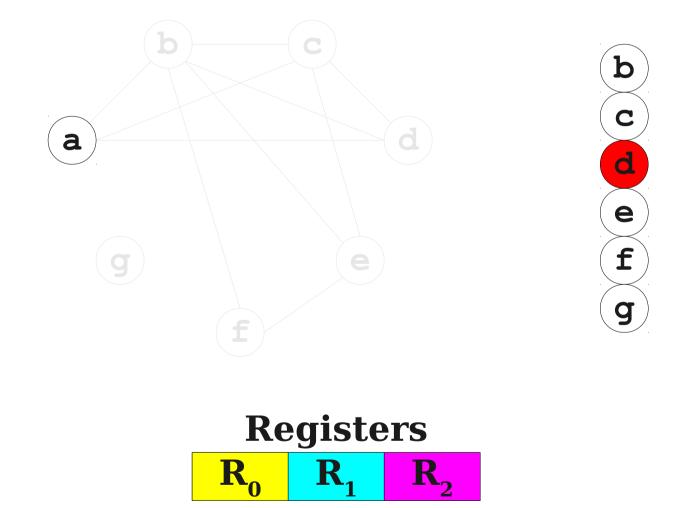


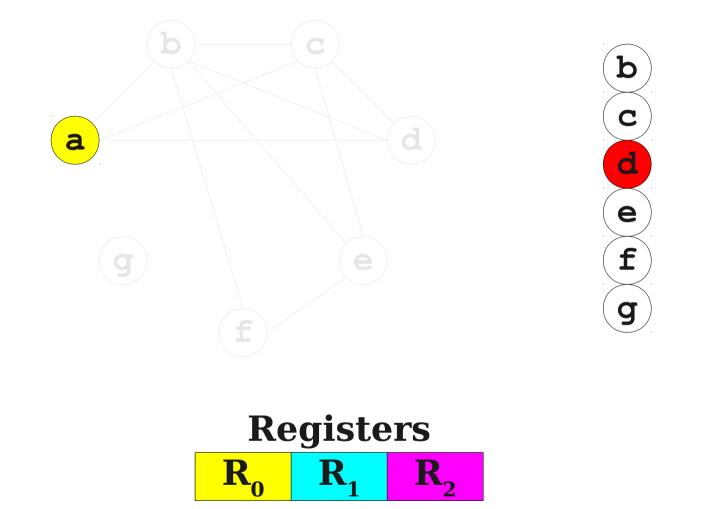


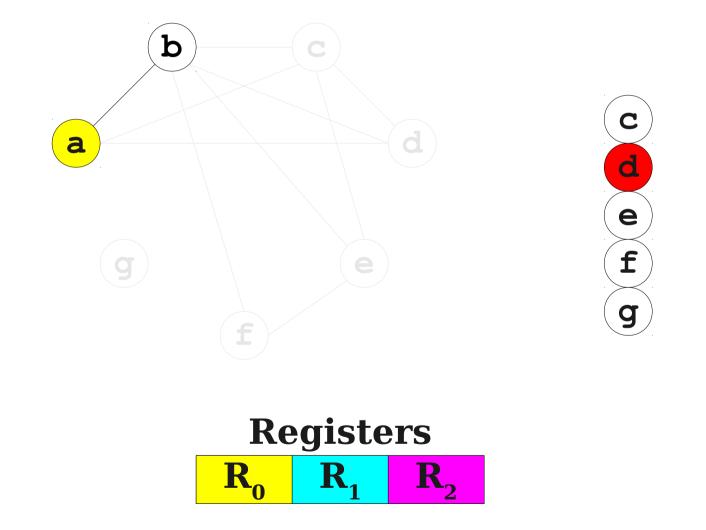


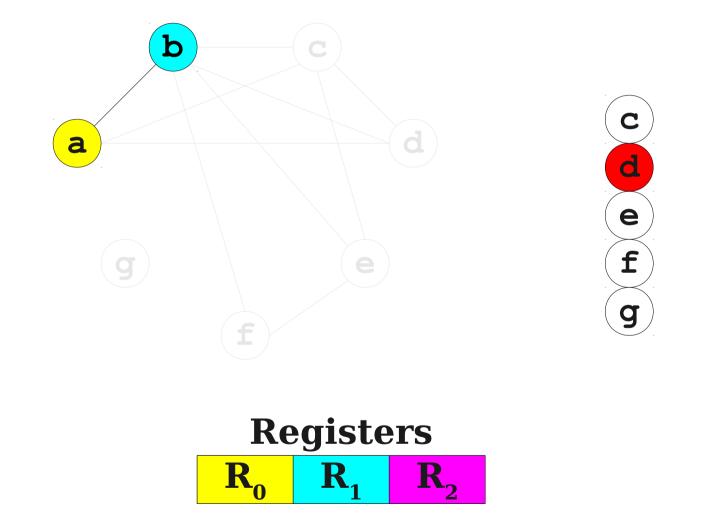


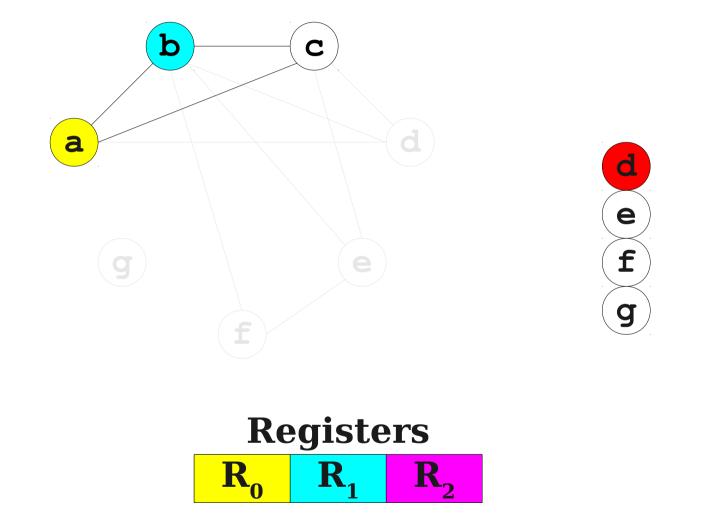


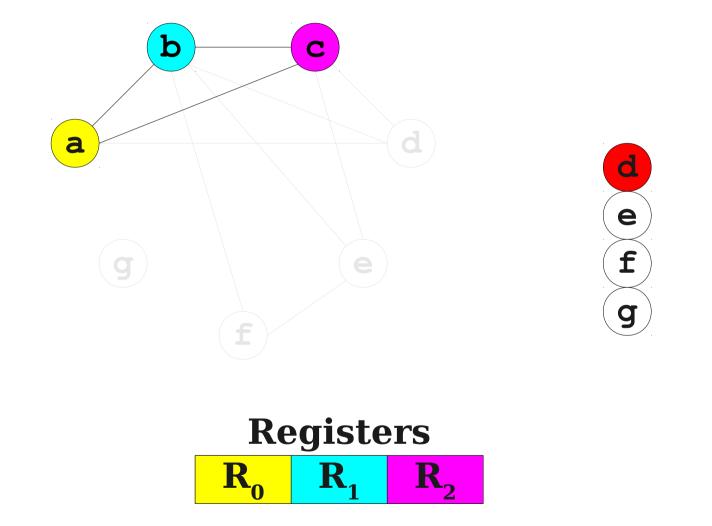


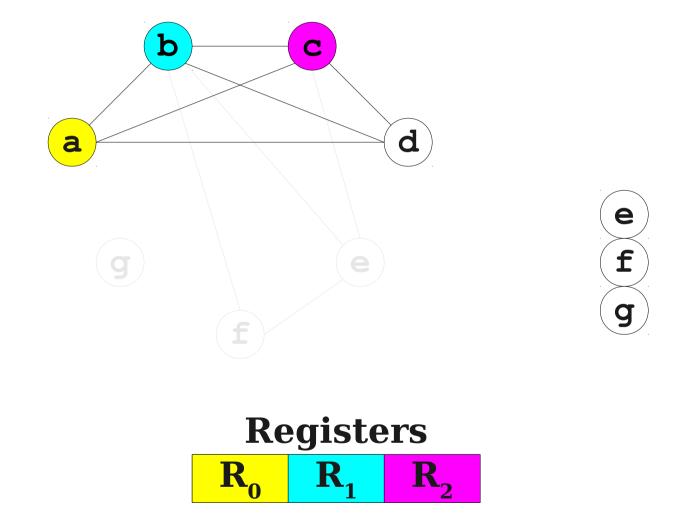


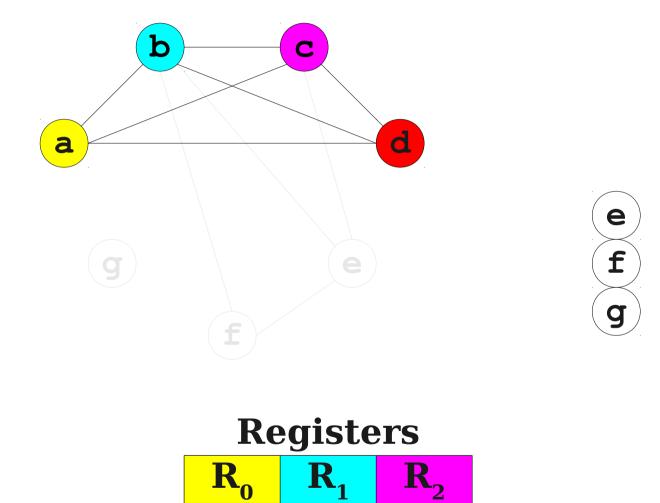


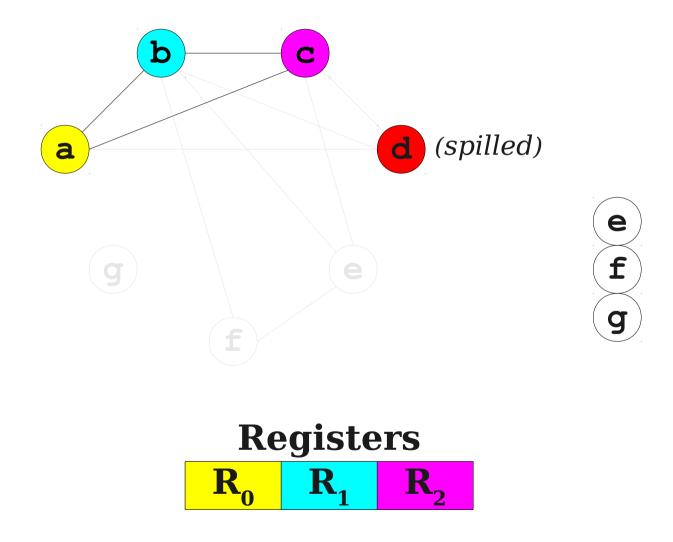


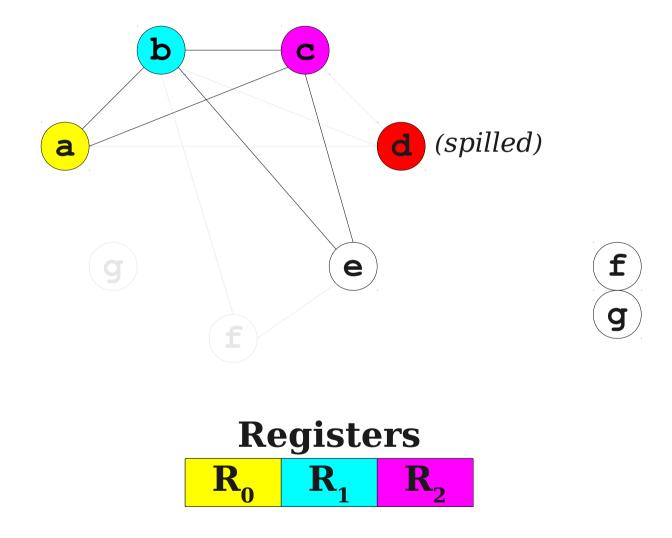


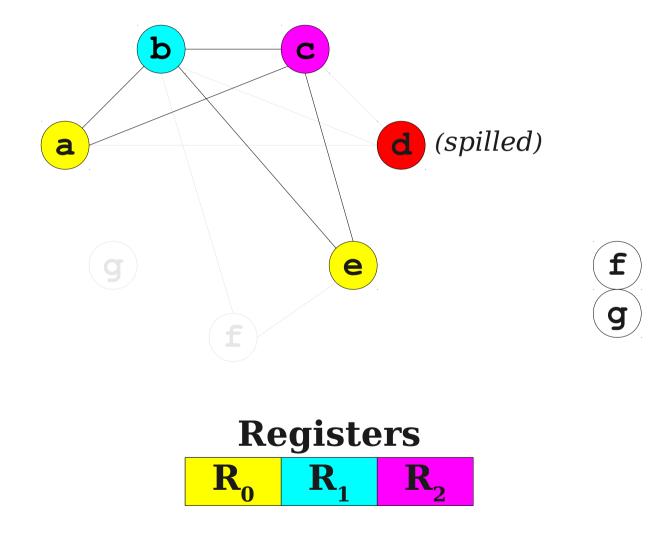


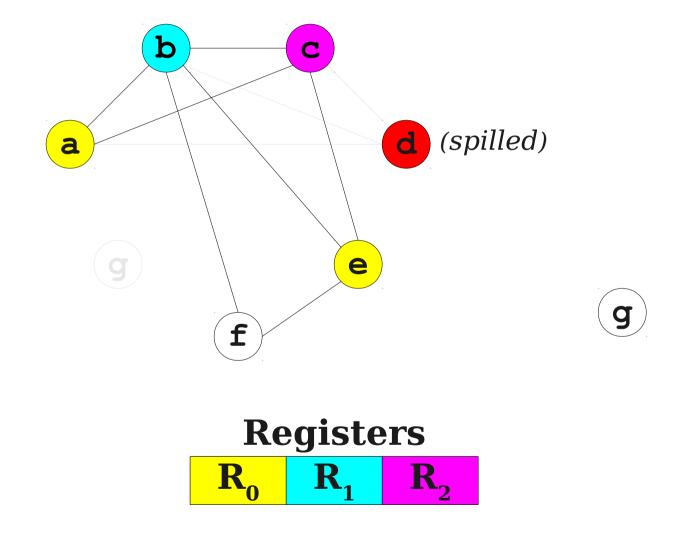


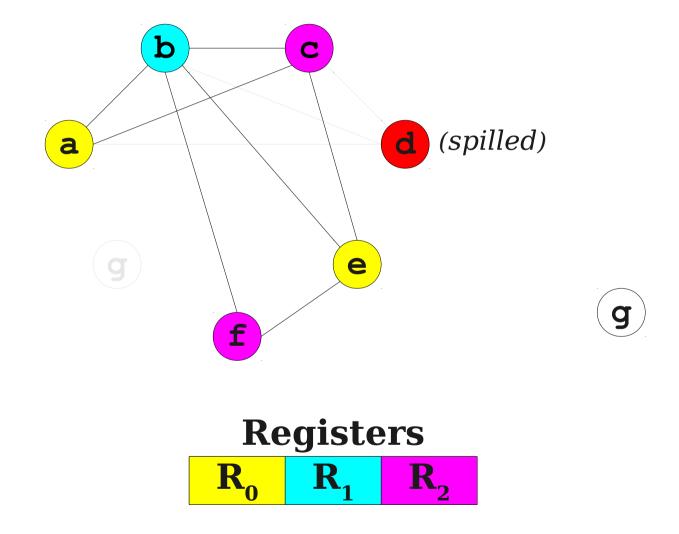


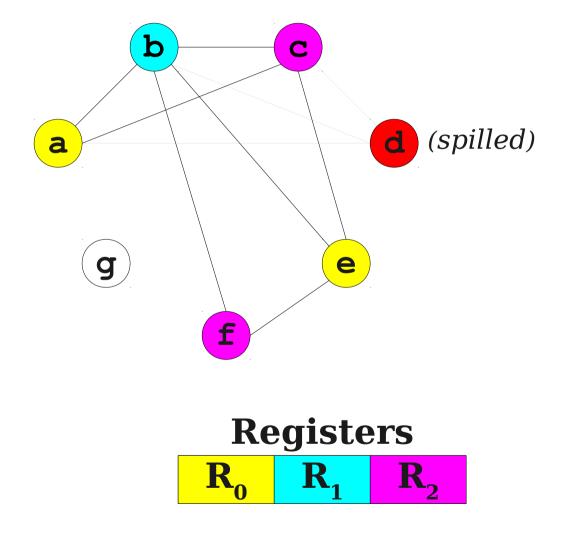


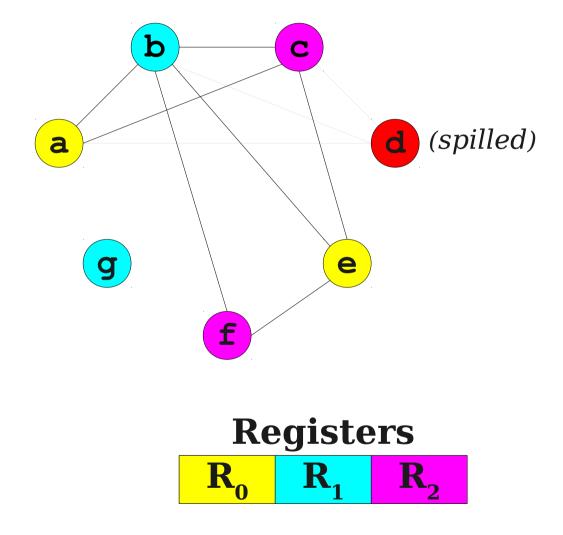


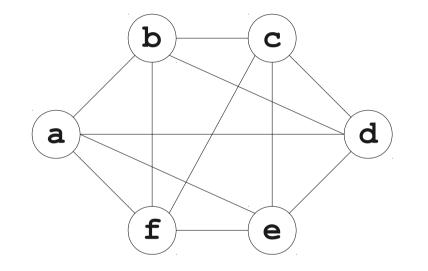


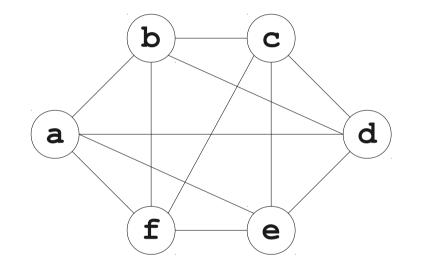




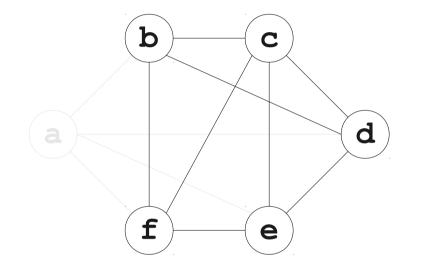






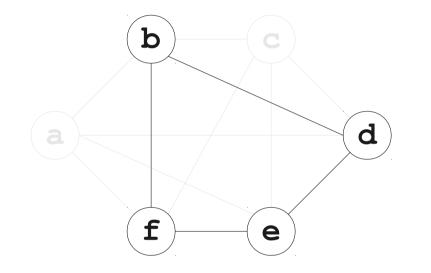


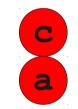




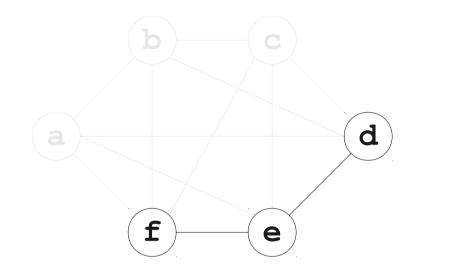


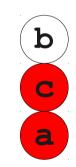




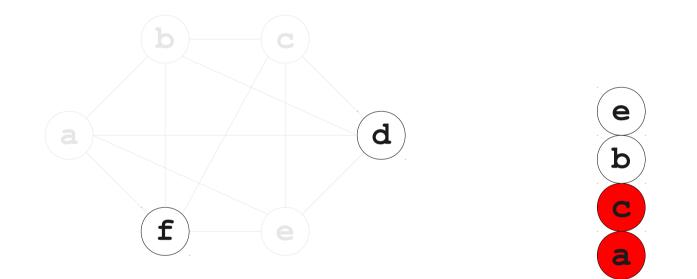




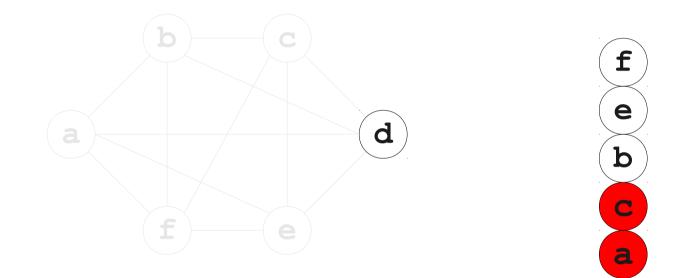




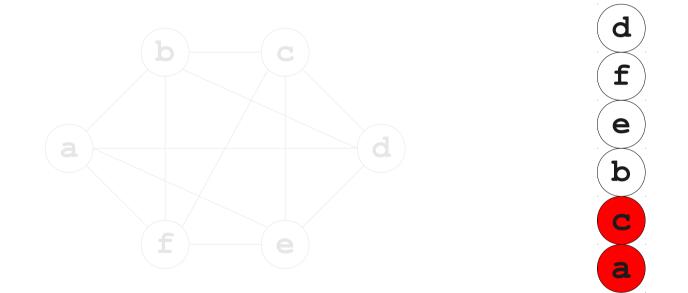




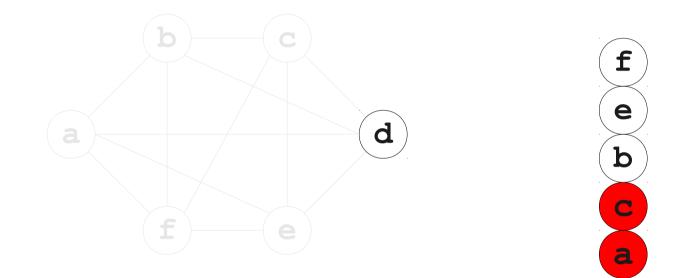




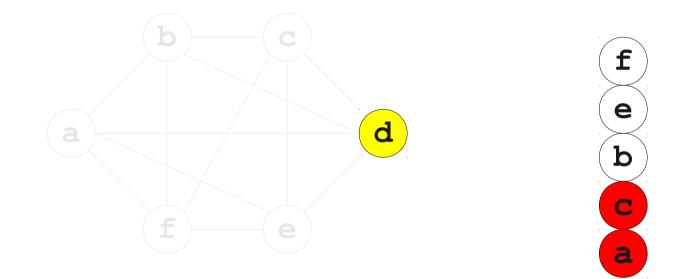




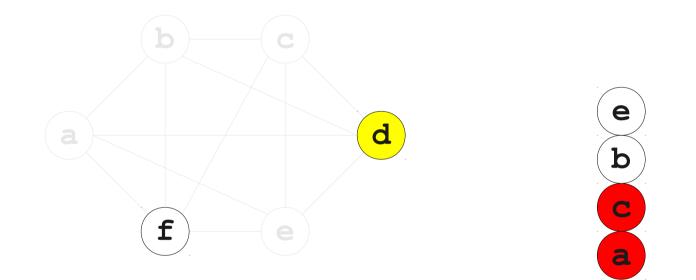




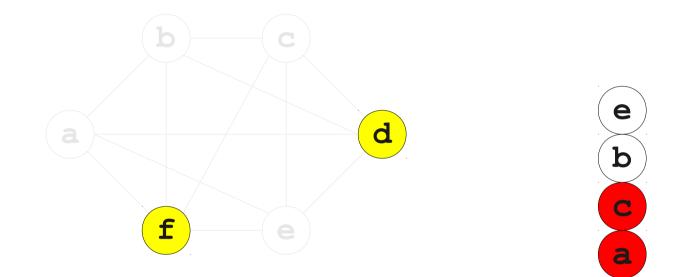




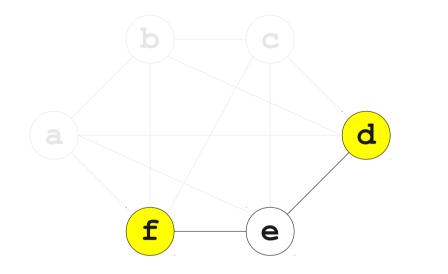


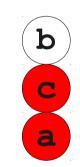




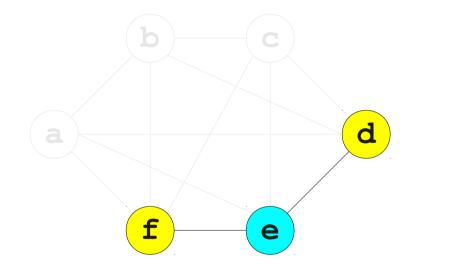


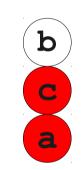




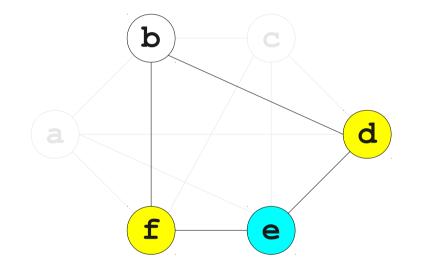


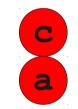




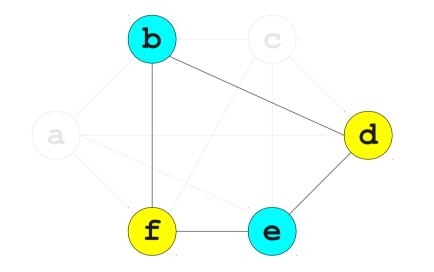


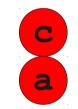




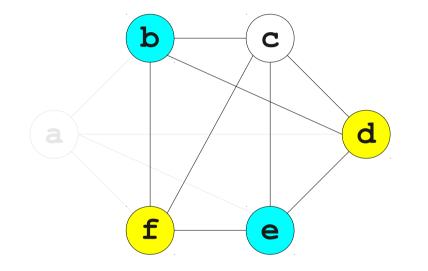






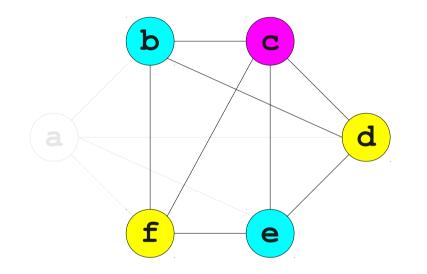






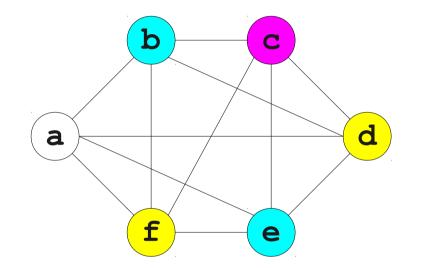




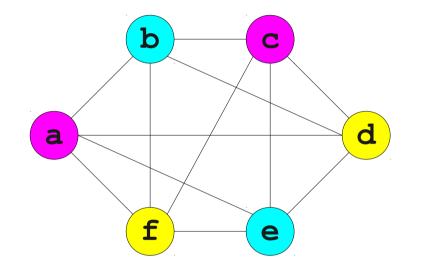




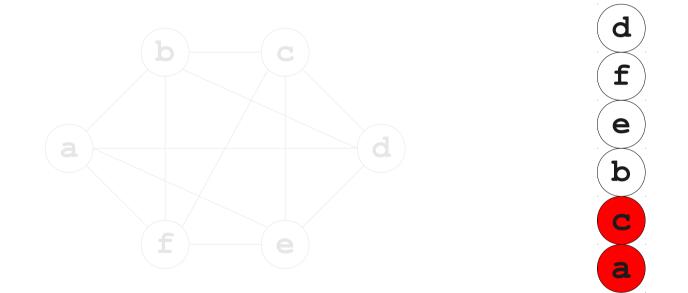




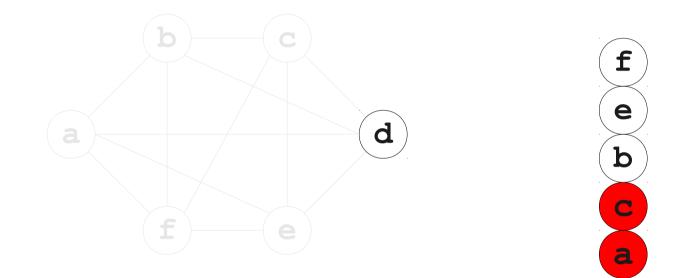




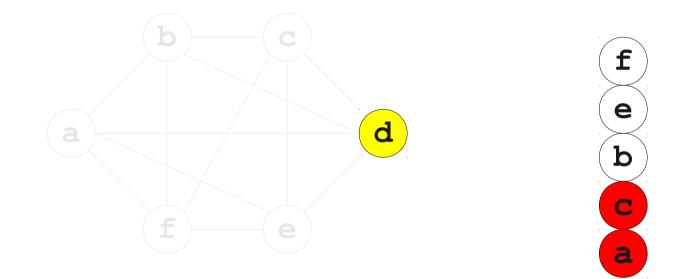




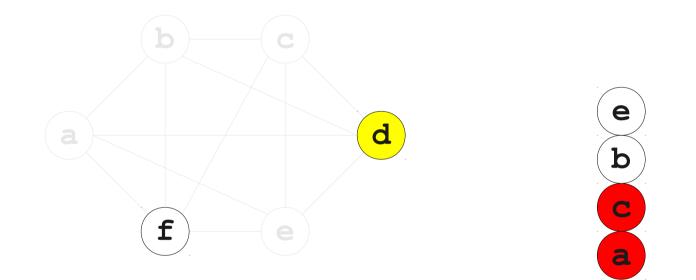




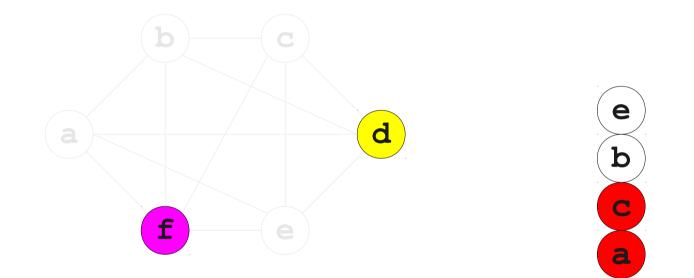




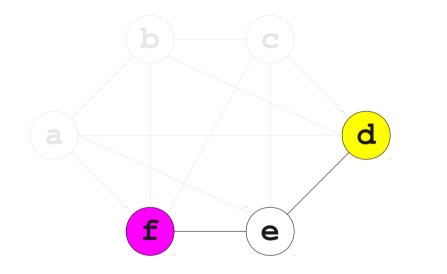


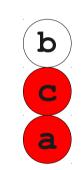




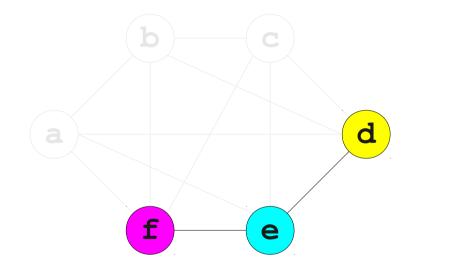


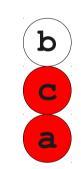




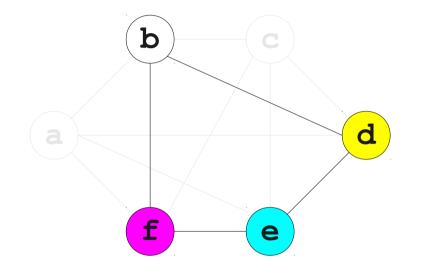


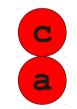




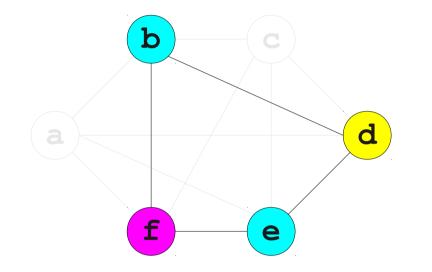


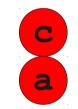




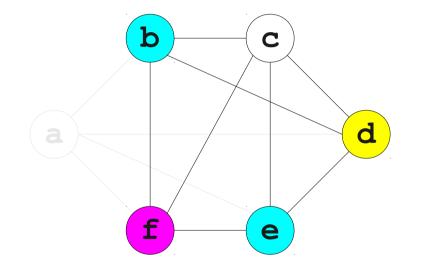






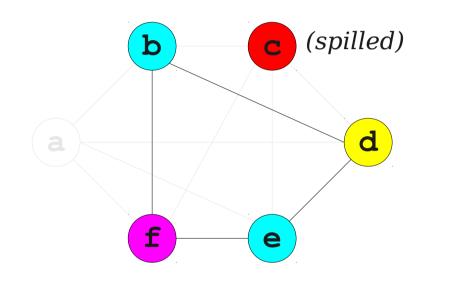






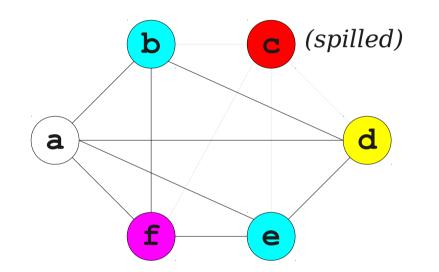




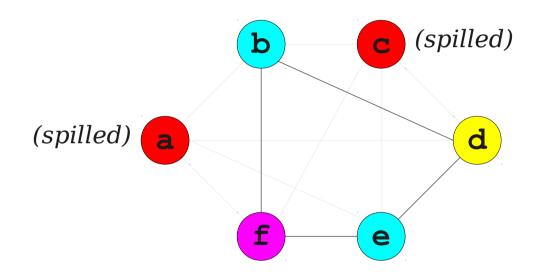














Chaitin's Algorithm

Chaitin's algorithm is efficient (O(|V| + |E|), simple to implement

Chaitin's Algorithm

Chaitin's algorithm is efficient (O(|V| + |E|), simple to implement

- How good the coloring is depends on the order we color the nodes to the graph
 - called the elimination ordering

Chaitin's Algorithm

Chaitin's algorithm is efficient (O(|V| + |E|), simple to implement

- How good the coloring is depends on the order we color the nodes to the graph
 - called the elimination ordering
- For every graph, there is a elimination ordering such that Chaitin's algorithm produces an optimal coloring
 - therefore finding this optimal elimination ordering for a general graph is NP-complete

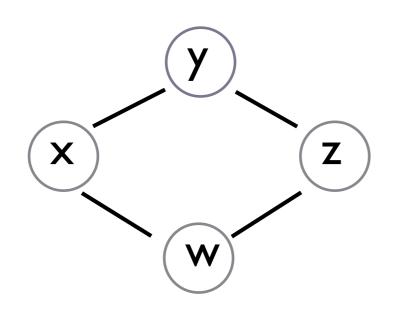
Graph Coloring SSA Programs

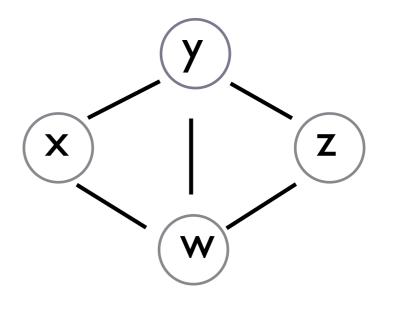
Hack et al, "Register Allocation for Programs in SSA-Form", Compiler Construction 2006

Graph Coloring SSA Programs

Hack et al, "Register Allocation for Programs in SSA-Form", Compiler Construction 2006

- The interference graphs of an SSA program are all chordal
 - Every cycle >= 4 nodes has a **chord**





Not chordal

chordal

Theorem: Every chordal graph has a **perfect** elimination ordering

Theorem: Every chordal graph has a **perfect** elimination ordering

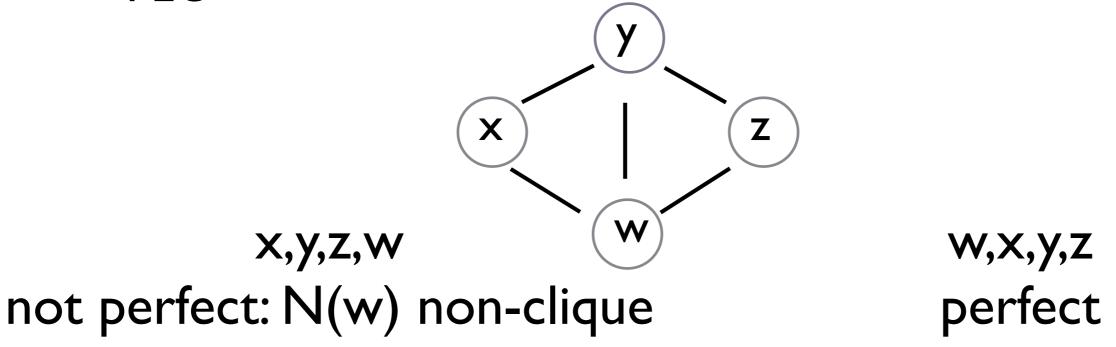
 a total ordering of nodes v1,v2,v3,... such that for each vi, vi forms a clique with all its neighbors earlier in the order

Theorem: Every chordal graph has a **perfect** elimination ordering

- a total ordering of nodes v1,v2,v3,... such that for each vi, vi forms a clique with all its neighbors earlier in the order
- Chaitin's algo produces an optimal coloring if we use a PEO

Theorem: Every chordal graph has a **perfect** elimination ordering

- a total ordering of nodes v1,v2,v3,... such that for each vi, vi forms a clique with all its neighbors earlier in the order
- Chaitin's algo produces an optimal coloring if we use a PEO



Every SSA Interference Graph is Chordal

Theorem: A graph is chordal iff it has a **perfect** elimination ordering

• a total ordering of nodes v1,v2,v3,... such that for each vi, vi forms a clique with all its neighbors later in the order

SSA programs have a simple PEO:

"in-scope" or "dominance" relation

Every SSA Interference Graph is Chordal

Theorem: A graph is chordal iff it has a **perfect** elimination ordering

• a total ordering of nodes v1,v2,v3,... such that for each vi, vi forms a clique with all its neighbors later in the order

SSA programs have a simple PEO:

"in-scope" or "dominance" relation

 a variable x dominates y if is in scope when y is defined (includes simultaneous defs)

Every SSA Interference Graph is Chordal

Theorem: A graph is chordal iff it has a **perfect** elimination ordering

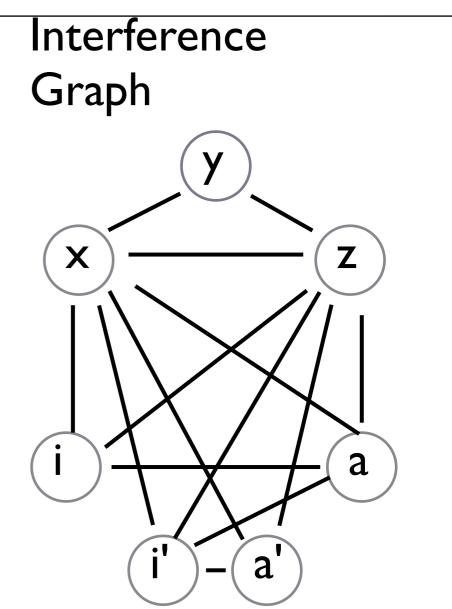
• a total ordering of nodes v1,v2,v3,... such that for each vi, vi forms a clique with all its neighbors later in the order

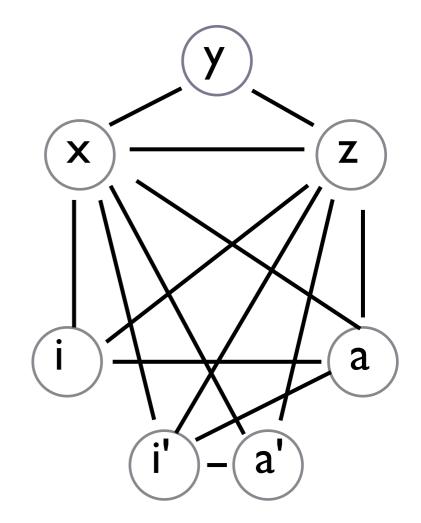
SSA programs have a simple PEO:

"in-scope" or "dominance" relation

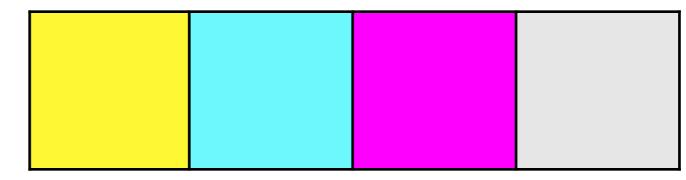
- a variable x dominates y if is in scope when y is defined (includes simultaneous defs)
- x's definition is "closer to the root" of the AST than y
- easy to compute: pre-order traversal of the nodes

```
def f(x,y,z):
  def loop(i,a):
    if i == 0:
      a * z
    else:
      let i' = i - 1 in
      let a' = a + x in
      icall(loop; i', a')
  end
  icall(loop; y, 0)
```

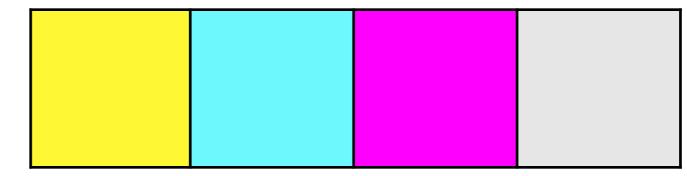




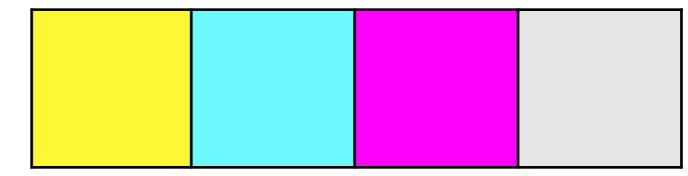
$$(\mathbf{x}, \mathbf{y}, \mathbf{z})$$
 (\mathbf{i}, \mathbf{a}) $(\mathbf{i}', \mathbf{a}')$



$$(\mathbf{x} \ \mathbf{y} \ \mathbf{z} \ \mathbf{i} \ \mathbf{a} \ \mathbf{i'} \ \mathbf{a'}$$

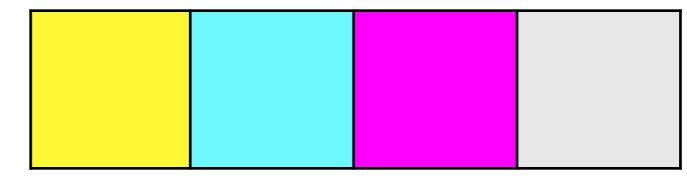


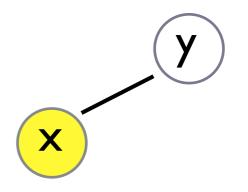
Χ



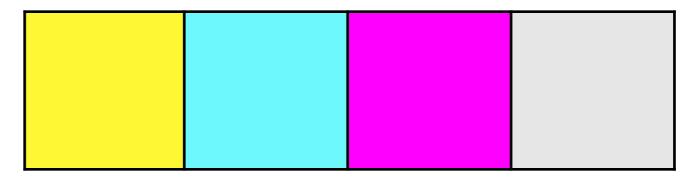
$$(x) (y) (z) (i) (a) (i') (a')$$

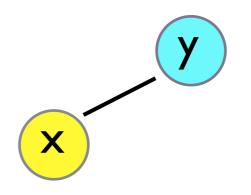
X



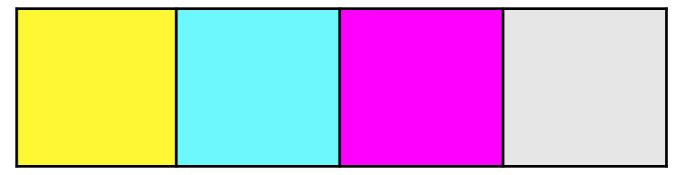


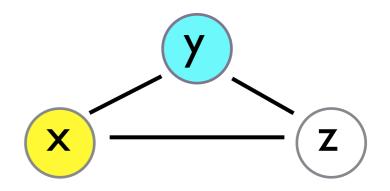
$$(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{i}, \mathbf{a}, \mathbf{i'}, \mathbf{a'})$$



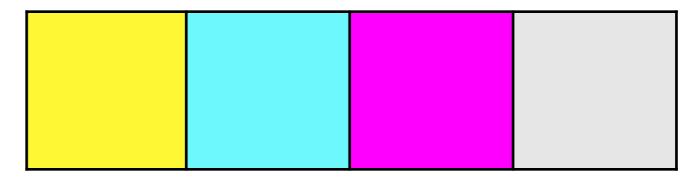


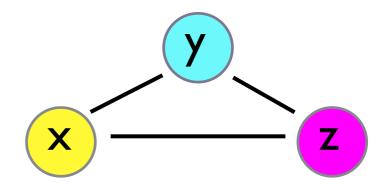
$$(\mathbf{x})(\mathbf{y})(\mathbf{z})(\mathbf{i})(\mathbf{a})(\mathbf{i'})(\mathbf{a'})$$



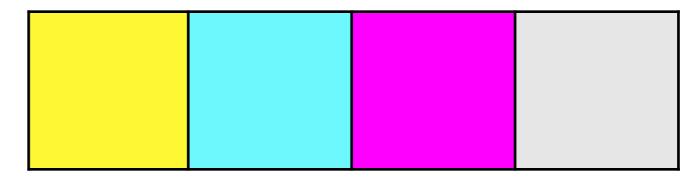


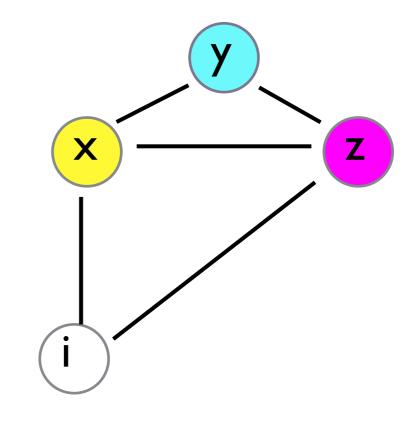
$$(\mathbf{x}, \mathbf{y}, \mathbf{z})$$
 (\mathbf{i}, \mathbf{a}) $(\mathbf{i}', \mathbf{a}')$



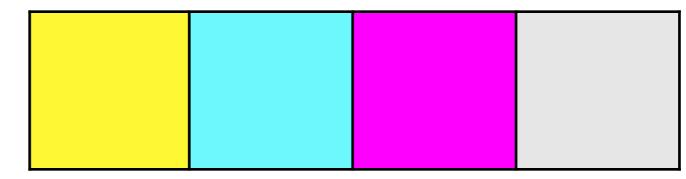


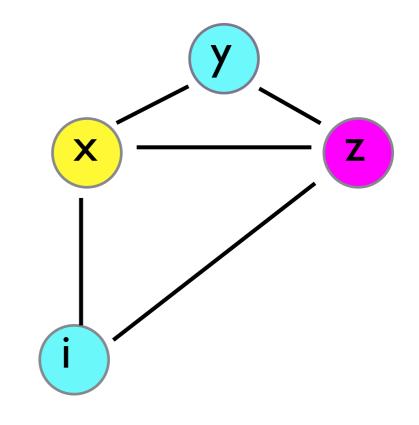
$$(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{i}, \mathbf{a}, \mathbf{i'}, \mathbf{a'})$$

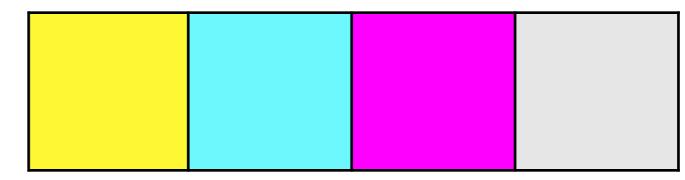


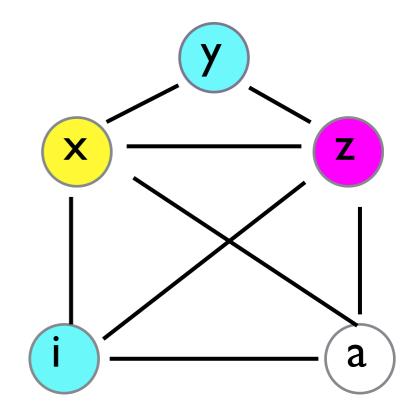


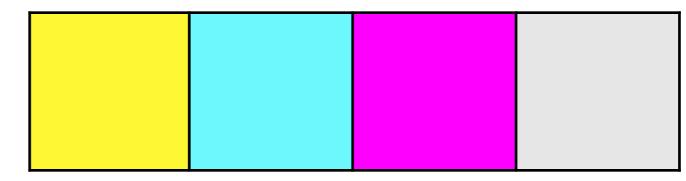
$$(\mathbf{x} \ \mathbf{y} \ \mathbf{z} \ \mathbf{i} \ \mathbf{a} \ \mathbf{i'} \ \mathbf{a'}$$

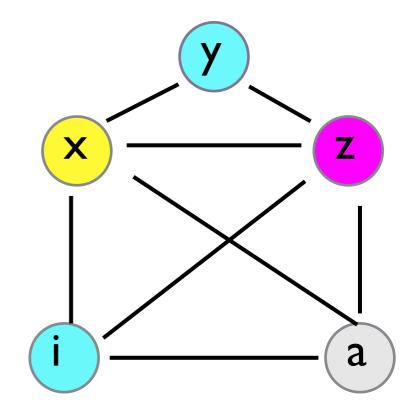


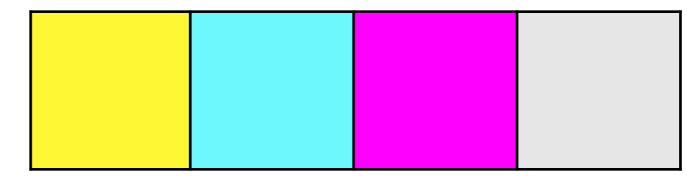


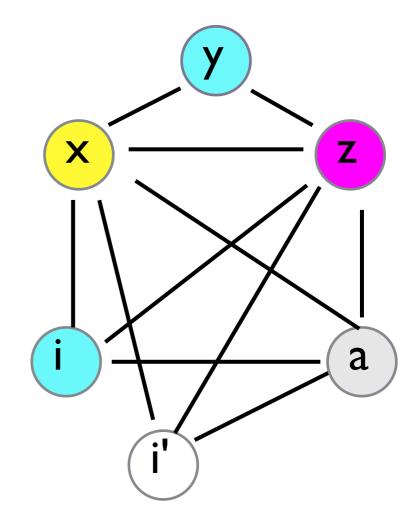


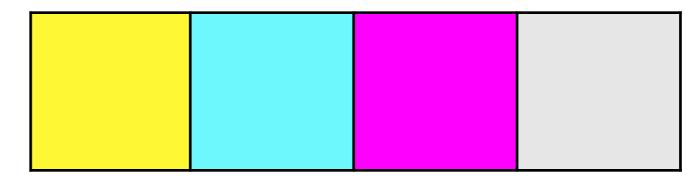


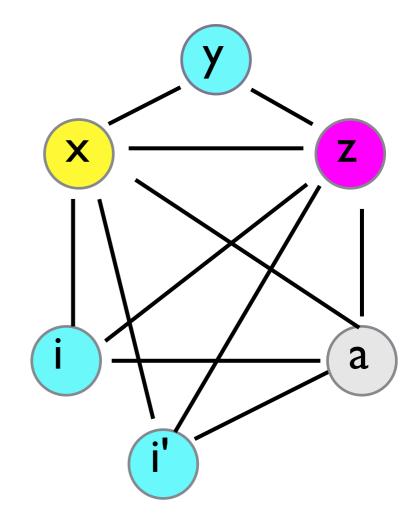




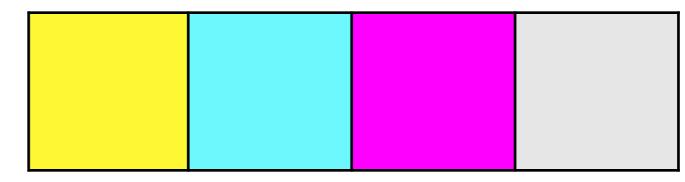


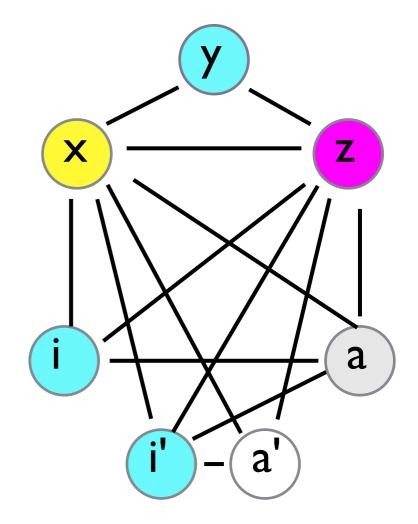




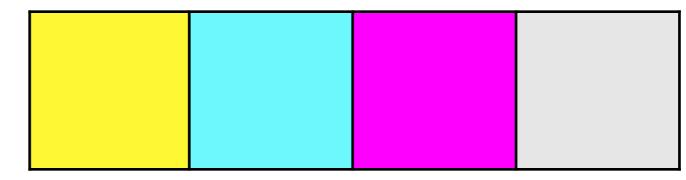


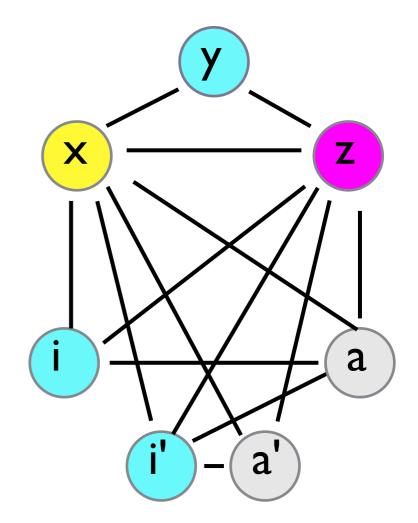
$$(\mathbf{x} \ \mathbf{y} \ \mathbf{z} \ \mathbf{i} \ \mathbf{a} \ \mathbf{i'} \ \mathbf{a'}$$



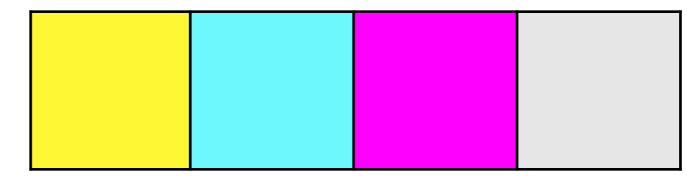


$$(\mathbf{x}, \mathbf{y}, \mathbf{z})$$
 (\mathbf{i}, \mathbf{a}) $(\mathbf{i}', \mathbf{a}')$





$$(\mathbf{x}, \mathbf{y}, \mathbf{z})$$
 (\mathbf{i}, \mathbf{a}) $(\mathbf{i}', \mathbf{a}')$



Using Register Assignment

I. For each function definition, we'll run liveness, conflict analysis and register allocation, producing a mapping from variable names to registers/stack offsets.

2. How does your code generation change?

Effects on Codegen

• No longer always put result in RAX: put result in

Want the result of y * z to go wherever x is stored, not RAX.

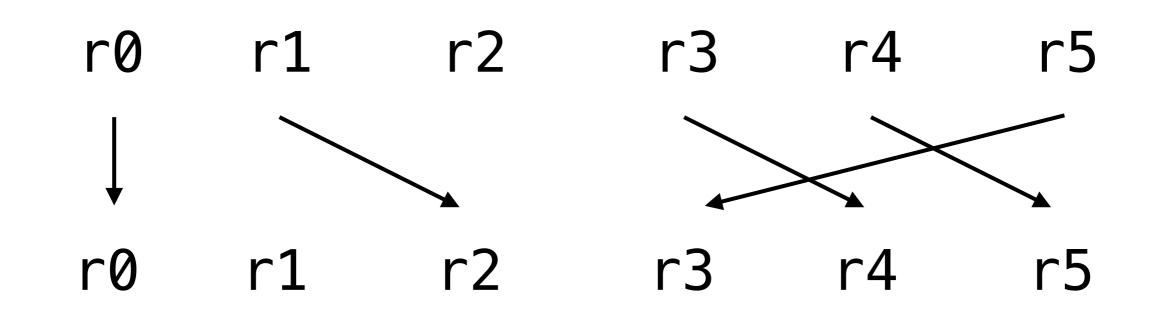
```
def f(a,b,c): e in
```

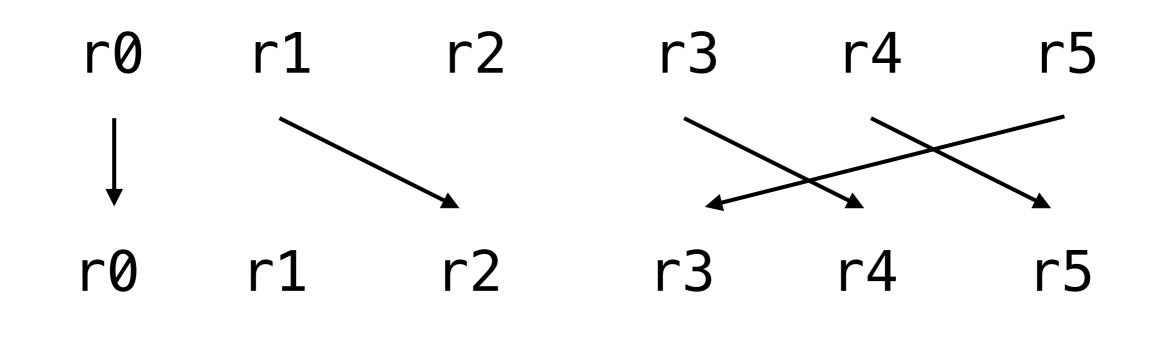
. . .

```
local_tail_call(f; [x,y,z])
```

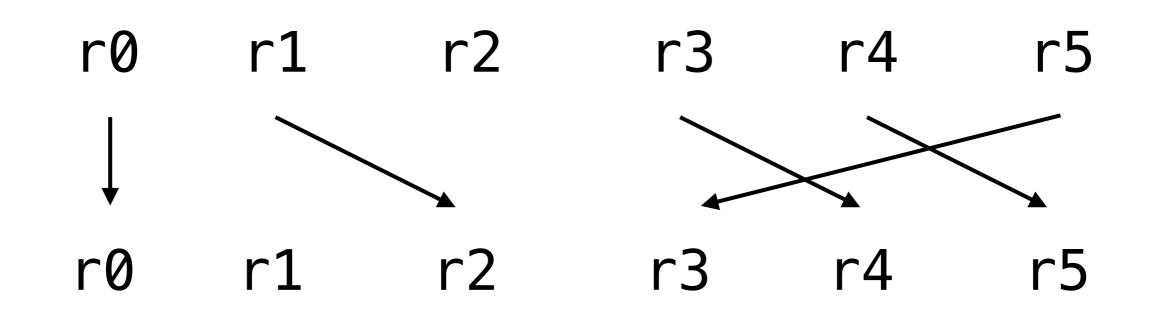
```
def f(a,b,c): e in
...
local_tail_call(f; [x,y,z])
```

mov r_x, r_a
mov r_y, r_b
mov r_z, r_z
jmp f

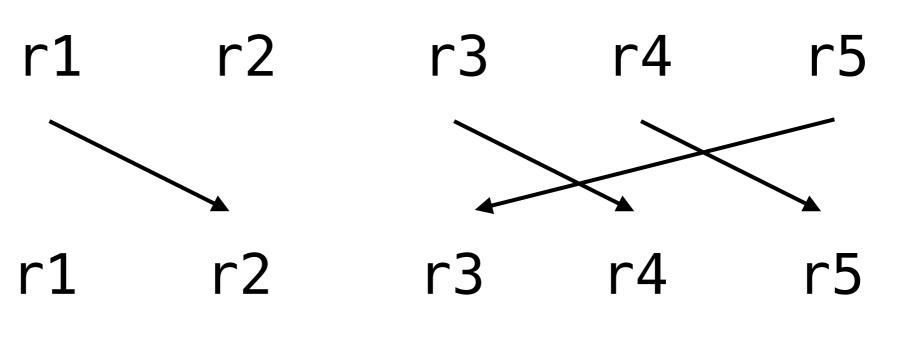




nop



nop mov r2, r1

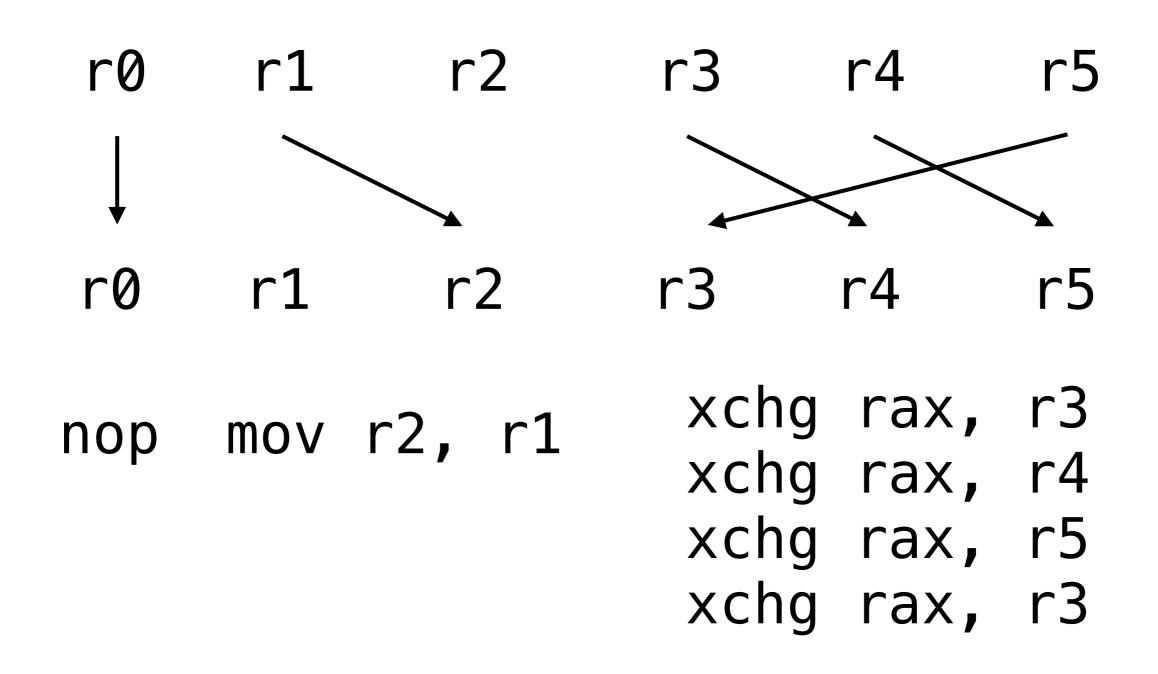


xchg rax, r3
xchg rax, r4
xchg rax, r5
xchg rax, r3

nop mov r2, r1

r0

r0



SSA reg allocation is polytime, but minimizing the resulting number of movs/xchg is NP hard

Now that we are using registers we need to take care to respect treatment of registers in the calling conventions we use.

Now that we are using registers we need to take care to respect treatment of registers in the calling conventions we use.

In System V AMD 64 Calling convention, registers are divided into two classes:

Now that we are using registers we need to take care to respect treatment of registers in the calling conventions we use.

In System V AMD 64 Calling convention, registers are divided into two classes:

• **volatile** aka **caller-save**: when you make a call, the value of these registers may change when the callee returns

Now that we are using registers we need to take care to respect treatment of registers in the calling conventions we use.

In System V AMD 64 Calling convention, registers are divided into two classes:

- **volatile** aka **caller-save**: when you make a call, the value of these registers may change when the callee returns
- non-volatile aka callee-save: when you make a call, the value of these registers will be the same when the callee returns

Volatile/Caller Save registers

volatile aka caller-save

let x = ... in let y = f(z) in x + y

if \mathbf{x} is stored in a volatile register, its value may be overwritten by the call.

Volatile/Caller Save registers

volatile aka caller-save

let x = ... in let y = f(z) in x + y

if \mathbf{x} is stored in a volatile register, its value may be overwritten by the call.

• Easy solution: save all live volatiles to the stack before a call, restore after the call

Volatile/Caller Save registers

volatile aka caller-save

let x = ... in let y = f(z) in x + y

if \mathbf{x} is stored in a volatile register, its value may be overwritten by the call.

- Easy solution: save all live volatiles to the stack before a call, restore after the call
- Harder solution: add nodes to interference graph for volatile registers, add conflicts at every non-tail call

Non-volatile/Callee Save registers

def f(x):

let $y = \dots$ in let z = x + y in z

if **y** is stored in a **non-volatile** register, its value must be **restored** when we return

Non-volatile/Callee Save registers

def f(x):

let $y = \dots$ in let z = x + y in z

if **y** is stored in a **non-volatile** register, its value must be **restored** when we return

• Easy solution: save all non-volatiles to the stack at the beginning of every global function def, restore them before every return/external tail call

Non-volatile/Callee Save registers

```
def f(x):
```

```
let y = \dots in
let z = x + y in z
```

if **y** is stored in a **non-volatile** register, its value must be **restored** when we return

- Easy solution: save all non-volatiles to the stack at the beginning of every global function def, restore them before every return/external tail call
- Harder solution: treat non-volatiles as "hidden args" of global fundefs, with ret/tail calls as uses.