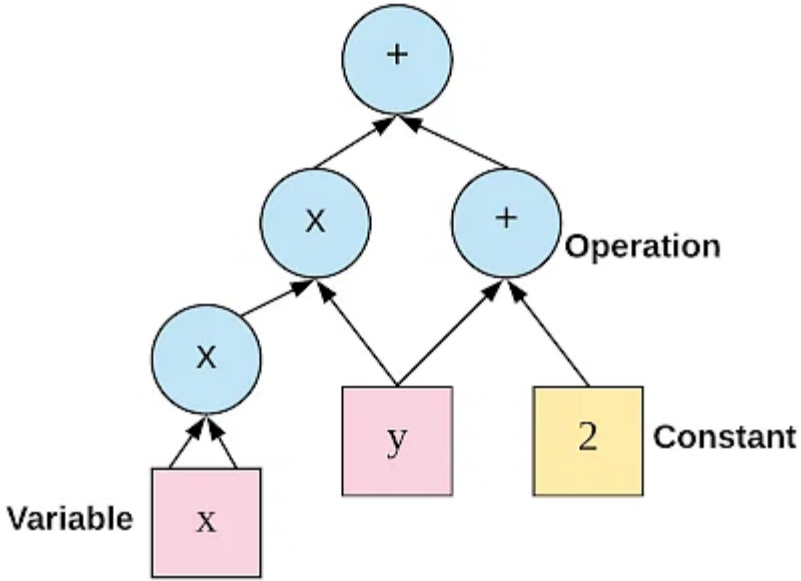
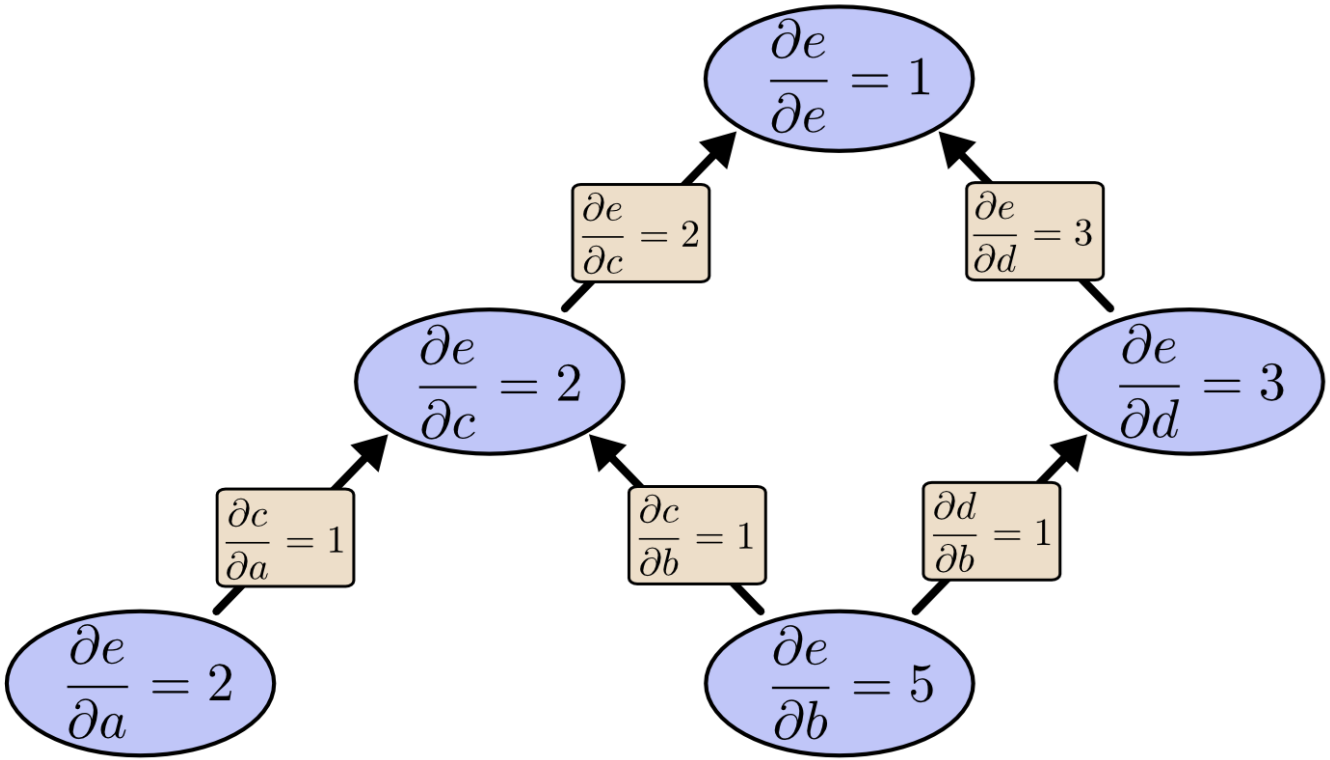
Dynamic Control Flow in Large-Scale Machine Learning

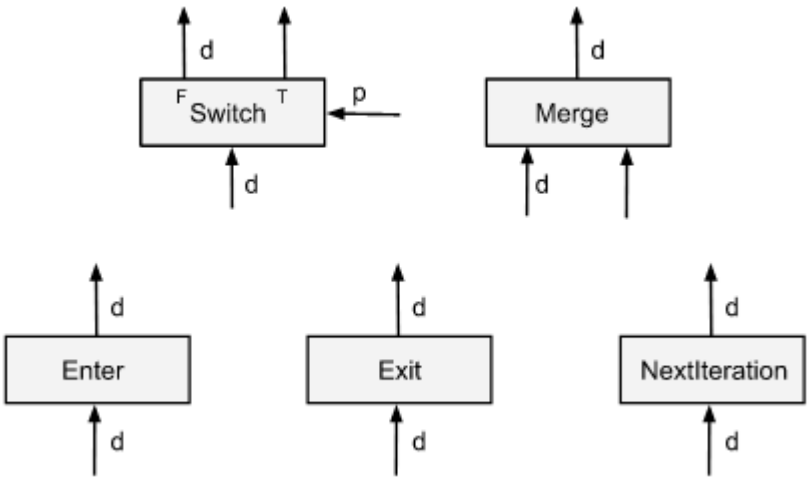
TensorFlow

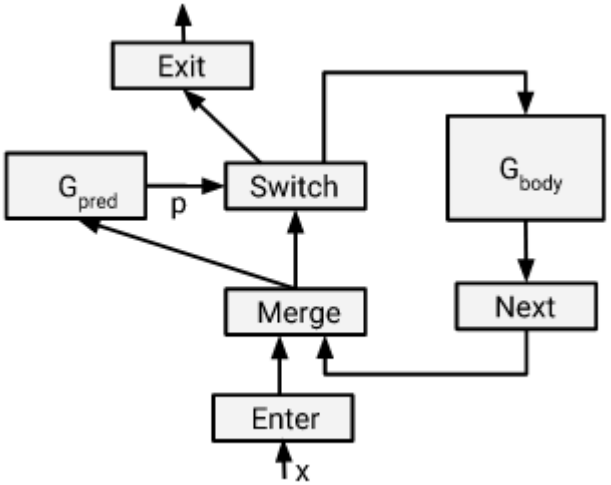
Computation graphs



Automatic differentiation



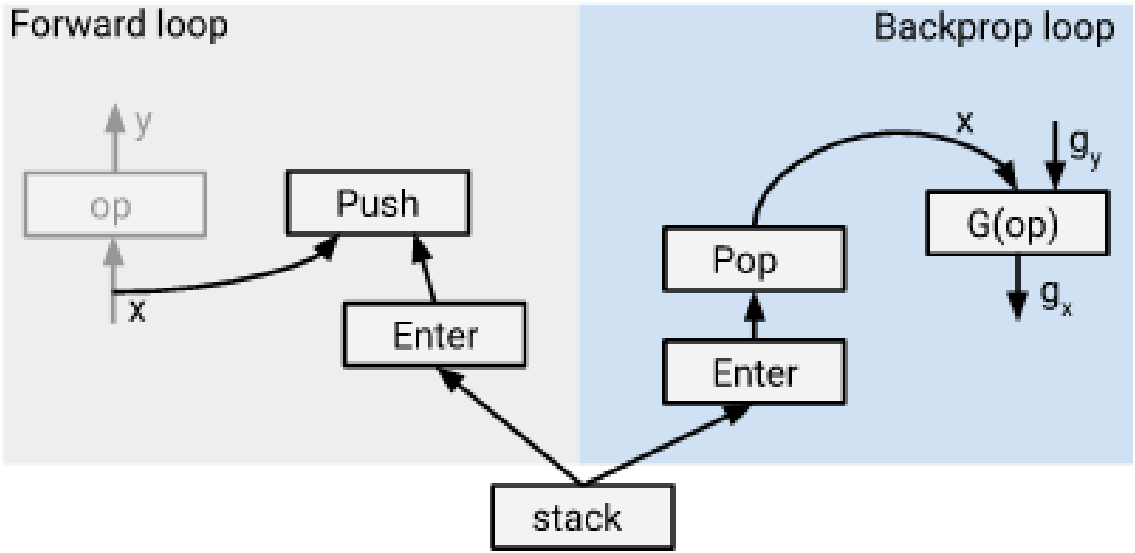

d
F Switch T
p
d
d
Merge
d
d
Enter
d
d
Exit
d
d
NextIteration
d
d

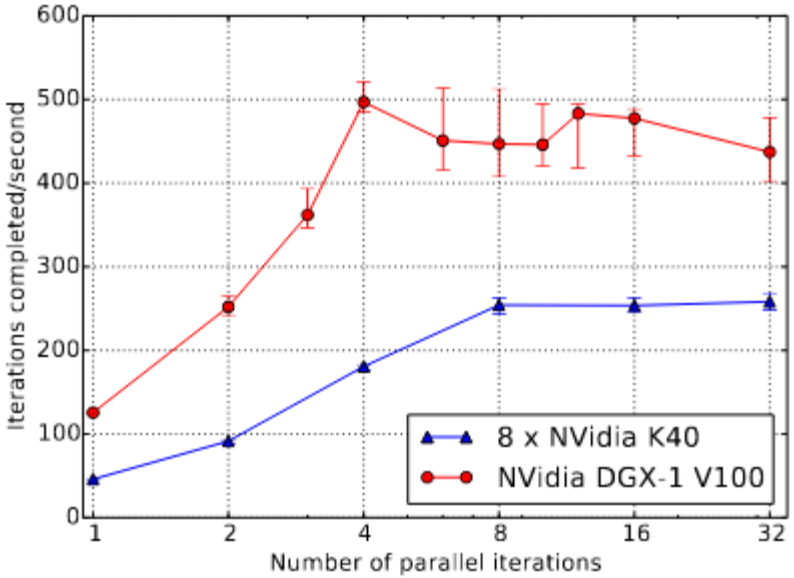

Exit
G pred
p
Switch
G body
Merge
Next
Enter
x

```
Eval(Switch(p, d), c) = (r1, r2), where
  r1 = (value(d), p || is_dead(d), tag(d))
  r2 = (value(d), !p || is_dead(d), tag(d))

Eval(Merge(d1, d2), c) = r, where
  r = if is_dead(d1) then d2 else d1

Eval(Enter(d, name), c) = r, where
  r = (value(d), is_dead(d), tag(d)/name/0)

Eval(Exit(d), c) = r, where
  r = (value(d), is_dead(d), c.parent.tag)

Eval(NextIteration(d), c) = r, where
  tag(d) = tag1/name/n
  r = (value(d), is_dead(d), tag1/name/(n+1))

Eval(Op(d1, ..., dm), c) = (r1, ..., rn), where
  value(r1), ..., value(rn) =
    Op(value(d1), ..., value(dm))
  is_dead(ri) =
    is_dead(d1) || ... || is_dead(dm), for all i
  tag(ri) = tag(d1), for all i
```

The gradient for `tf.cond(pred, true_fn, false_fn)` with output gradients `g_z` is

`tf.cond(pred, true_fn_grad(g_z), false_fn_grad(g_z))`.


Forward loop
Backprop loop
y
op
x
Push
Enter
stack
Pop
Enter
x
g_y
G(op)
g_x



Swap	Training time per loop iteration (ms), by sequence length						
	100	200	500	600	700	900	1000
Disabled	5.81	5.78	5.75	OOM	OOM	OOM	OOM
Enabled	5.76	5.76	5.73	5.72	5.77	5.74	5.74

Table 1: Training time per loop iteration for an LSTM model with increasing sequence lengths.