

This notebook is designed to show the problem with interchip communication on nengo-loihi

If any of the cells time out, interrupt the cell and start the notebook over.

Start out by importing necessary modules

```
In [1]: from tensorflow.keras.layers import *
from nengo_loihi.hardware.allocators import PartitionInterchip
import tensorflow as tf
import nengo_dl
import nengo
import numpy as np
import nengo_loihi
import sys
```

```
2022-01-10 11:48:22.048676: W tensorflow/stream_executor/platform/default/dso_loader.cc:64] Could not load dynamic library 'libcudart.so.11.0'; dLError: libcudart.so.11.0: cannot open shared object file: No such file or directory; LD_LIBRARY_PATH: /slurm/intel-archi/lib
2022-01-10 11:48:22.048717: I tensorflow/stream_executor/cuda/cudart_stub.cc:29] Ignore above cudart dLError if you do not have a GPU set up on your machine.
2022-01-10 11:48:26.167570: W tensorflow/stream_executor/platform/default/dso_loader.cc:64] Could not load dynamic library 'libcudart.so.11.0'; dLError: libcudart.so.11.0: cannot open shared object file: No such file or directory; LD_LIBRARY_PATH: /slurm/intel-archi/lib
2022-01-10 11:48:28.384400: W tensorflow/stream_executor/platform/default/dso_loader.cc:64] Could not load dynamic library 'libcuda.so.1'; dLError: libcuda.so.1: cannot open shared object file: No such file or directory; LD_LIBRARY_PATH: /slurm/intel-archi/lib
2022-01-10 11:48:28.384435: W tensorflow/stream_executor/cuda/cuda_driver.cc:269] failed call to cuInit: UNKNOWN ERROR (303)
```

Construct a simple convolutional network

```
In [2]: # define a convolutional network
def simple_neural_network():
    inputs = Input(shape=(64, 64, 3))
    spiking_input = Activation(tf.nn.elu)(inputs) # this will be replaced by SpikingRectifiedLinear
    conv_out1 = Conv2D(32, (5, 5), strides=(2, 2), padding='valid', activation=tf.nn.relu)
    conv_out2 = Conv2D(64, (3, 3), strides=(2, 2), padding='valid', activation=tf.nn.relu)
    conv_out3 = Conv2D(128, (2, 2), strides=(2, 2), padding='valid', activation=tf.nn.relu)
    conv_out4 = Conv2D(256, (2, 2), strides=(2, 2), padding='valid', activation=tf.nn.relu)

    flat_out = Flatten()(conv_out4)
    output = Dense(4, activation=None, name="dense", use_bias = False)(flat_out)

    model = tf.keras.Model(inputs=inputs, outputs=output)
    return model
```

Convert network to spiking implementation

```
In [3]: ann_model = simple_neural_network()
nengo_converter = nengo_dl.Converter(
    ann_model,
    swap_activations={tf.nn.relu: nengo_loihi.LoihiSpikingRectifiedLinear(), # this is c
```

```
tf.nn.elu: nengo.SpikingRectifiedLinear()),
```

```
2022-01-10 11:48:29.816575: W tensorflow/stream_executor/platform/default/dso_loader.cc:64] Could not load dynamic library 'libcuda.so.1'; dlerror: libcuda.so.1: cannot open shared object file: No such file or directory; LD_LIBRARY_PATH: /slurm/intel-archi/lib
2022-01-10 11:48:29.816614: W tensorflow/stream_executor/cuda/cuda_driver.cc:269] failed call to cuInit: UNKNOWN ERROR (303)
2022-01-10 11:48:29.816640: I tensorflow/stream_executor/cuda/cuda_diagnostics.cc:156] kernel driver does not appear to be running on this host (ncl-edu): /proc/driver/nvidia/version does not exist
```

Specify first layer as running off chip

In [4]:

```
with nengo_converter.net as net:
    nengo_loihi.add_params(net) # allow on_chip to be set
    net.config[nengo_converter.layers[ann_model.layers[1]].ensemble].on_chip = False
```

Define our input to the SNN

This is fake imagery data with extremely low spike rates.

In [5]:

```
nengo_input = nengo_converter.inputs[ann_model.layers[0]]
with nengo_converter.net as net:
    nengo_input.output = nengo.processes.PresentInput(
        np.random.random((10, 64, 64, 3)), presentation_time=.25
    )
```

Specify our block sizes

If a block shape of (n, m, c) is specified, it means that for that specific layer output of the neural network it will try to allocate neurons with this block shape. The following inequality holds due to hardware constraints $n \cdot m \cdot c \leq 1024$

By decreasing the dimensions of n, m, or c we can force our neural network to be partitioned inefficiently and use more cores.

In the next cell, I use efficient block shapes so you can see that the neural network successfully functions on 1 chip.

In [6]:

```
block_sizes = [None, (16, 16, 4), (8, 8, 16), None]
conv_layers = [layer for layer in ann_model.layers if "conv" in str(layer).lower()]
for layer, block_size, layer_idx in zip(conv_layers, block_sizes, range(len(conv_layers))):
    if block_size == None: #
        continue
    output_shape = tuple(layer.output.shape[1:])
    with nengo_converter.net as net:
        net.config[net.ensembles[layer_idx+1]].block_shape = nengo_loihi.BlockShape(block_size)
```

Try Inference on the network for a set number of epochs

If this cell fails with an srun timeout, restart it until it works.

In [7]:

```
nengo_loihi.hardware.interface.HostSnip.recv_timeout = 10.0 # Default is 0.01 (10 ms)
nengo_loihi.hardware.interface.HostSnip.recv_retries = 100 # Default is 10
with nengo_loihi.Simulator(net, remove_passthrough=False, precompute = False, target="low"):
    print(f"Precompute = {loihi_sim.precompute}")
    # iterate over the timesteps

    for i in range(10): # iterate over the batches
        loihi_sim.clear_probes() # prevents the buffer from filling up and crashing our
        loihi_sim.run(.25)
```

/homes/mjurado3/miniconda3/envs/loihi_vishal/lib/python3.9/site-packages/nengo_loihi/simulator.py:159: UserWarning: Model is precomputable. Setting precompute=False may slow execution.

```
warnings.warn(
```

```
INFO:DRV: SLURM is being run in background
INFO:DRV: Connecting to 10.212.98.110:38283
INFO:DRV: Host server up.....Done 0.20s
INFO:DRV: Encoding axons/synapses.....Done 12.00s
INFO:DRV: Compiling Embedded snips....Done 1.62s
INFO:DRV: Compiling Host snips.....Done 1.18s
INFO:DRV: Compiling MPDS Registers....Done 0.67ms
INFO:HST: Args chip=0 cpu=0 /homes/mjurado3/miniconda3/envs/loihi_vishal/lib/python3.9/
site-packages/nxsdk/driver/compiler/../../../../temp/1641844162.7226791/launcher_chip0_lmt
0.bin --chips=1 --remote-relay=1
INFO:HST: Nx...
INFO:DRV: Booting up.....Done 0.65s
INFO:DRV: Encoding probes.....Done 0.53ms
Precompute = False
INFO:DRV: Transferring probes.....Done 5.66ms
INFO:DRV: Configuring registers.....Done 2.58s
INFO:DRV: Transferring spikes.....Done 0.56ms
INFO:HST: [Host] Listening for client
INFO:HST: [Host] Connected to client
INFO:HST: chip=0 cpu=0 time 100
INFO:HST: chip=0 cpu=0 time 200
INFO:HST: chip=0 cpu=0 Waited to exit (nonsense sum -13580)
INFO:DRV: Executing.....Done 0.45s
INFO:DRV: Processing timeseries.....Done 1.06ms
INFO:DRV: Transferring probes.....Done 1.08ms
INFO:DRV: Configuring registers.....Done 3.65ms
INFO:DRV: Transferring spikes.....Done 1.06ms
INFO:HST: chip=0 cpu=0 time 300
INFO:HST: chip=0 cpu=0 time 400
INFO:HST: chip=0 cpu=0 time 500
INFO:HST: chip=0 cpu=0 Waited to exit (nonsense sum -13580)
INFO:DRV: Executing.....Done 0.45s
INFO:DRV: Processing timeseries.....Done 1.88ms
INFO:DRV: Transferring probes.....Done 0.90ms
INFO:DRV: Configuring registers.....Done 5.57ms
INFO:DRV: Transferring spikes.....Done 0.61ms
INFO:HST: chip=0 cpu=0 time 600
INFO:HST: chip=0 cpu=0 time 700
INFO:DRV: Executing.....Done 0.45s
INFO:HST: chip=0 cpu=0 Waited to exit (nonsense sum -13580)
INFO:DRV: Processing timeseries.....Done 0.57ms
INFO:DRV: Transferring probes.....Done 0.54ms
INFO:DRV: Configuring registers.....Done 3.83ms
INFO:DRV: Transferring spikes.....Done 0.53ms
INFO:HST: chip=0 cpu=0 time 800
INFO:HST: chip=0 cpu=0 time 900
INFO:HST: chip=0 cpu=0 time 1000
INFO:HST: chip=0 cpu=0 Waited to exit (nonsense sum -13580)
INFO:DRV: Executing.....Done 0.45s
INFO:DRV: Processing timeseries.....Done 0.60ms
INFO:DRV: Transferring probes.....Done 1.43ms
INFO:DRV: Configuring registers.....Done 5.33ms
INFO:DRV: Transferring spikes.....Done 0.67ms
INFO:HST: chip=0 cpu=0 time 1100
INFO:HST: chip=0 cpu=0 time 1200
INFO:HST: chip=0 cpu=0 Waited to exit (nonsense sum -13580)
INFO:DRV: Executing.....Done 0.45s
INFO:DRV: Processing timeseries.....Done 0.74ms
INFO:DRV: Transferring probes.....Done 0.72ms
INFO:DRV: Configuring registers.....Done 3.35ms
INFO:DRV: Transferring spikes.....Done 0.62ms
INFO:HST: chip=0 cpu=0 time 1300
INFO:HST: chip=0 cpu=0 time 1400
INFO:HST: chip=0 cpu=0 time 1500
INFO:DRV: Executing.....Done 0.45s
```

```

INFO:HST: chip=0 cpu=0 Waited to exit (nonsense sum -13580)
INFO:DRV: Processing timeseries.....Done 0.60ms
INFO:DRV: Transferring probes.....Done 0.57ms
INFO:DRV: Configuring registers.....Done 3.88ms
INFO:DRV: Transferring spikes.....Done 1.16ms
INFO:HST: chip=0 cpu=0 time 1600
INFO:HST: chip=0 cpu=0 time 1700
INFO:DRV: Executing.....Done 0.45s
INFO:HST: chip=0 cpu=0 Waited to exit (nonsense sum -13580)
INFO:DRV: Processing timeseries.....Done 0.95ms
INFO:DRV: Transferring probes.....Done 0.51ms
INFO:DRV: Configuring registers.....Done 0.01s
INFO:DRV: Transferring spikes.....Done 0.58ms
INFO:HST: chip=0 cpu=0 time 1800
INFO:HST: chip=0 cpu=0 time 1900
INFO:HST: chip=0 cpu=0 time 2000
INFO:HST: chip=0 cpu=0 Waited to exit (nonsense sum -13580)
INFO:DRV: Executing.....Done 0.45s
INFO:DRV: Processing timeseries.....Done 2.18ms
INFO:DRV: Transferring probes.....Done 0.58ms
INFO:DRV: Configuring registers.....Done 0.01s
INFO:DRV: Transferring spikes.....Done 2.98ms
INFO:HST: chip=0 cpu=0 time 2100
INFO:HST: chip=0 cpu=0 time 2200
INFO:HST: chip=0 cpu=0 Waited to exit (nonsense sum -13580)
INFO:DRV: Executing.....Done 0.45s
INFO:DRV: Processing timeseries.....Done 1.93ms
INFO:DRV: Transferring probes.....Done 0.96ms
INFO:DRV: Configuring registers.....Done 5.55ms
INFO:DRV: Transferring spikes.....Done 0.70ms
INFO:HST: chip=0 cpu=0 time 2300
INFO:HST: chip=0 cpu=0 time 2400
INFO:HST: chip=0 cpu=0 time 2500
INFO:HST: chip=0 cpu=0 Waited to exit (nonsense sum -13580)
INFO:DRV: Executing.....Done 0.45s
INFO:DRV: Processing timeseries.....Done 0.81ms
INFO:HST: [Host] Received shutdown signal: -1
INFO:HST: [Host] Wrote superhost shutdown signal: 8192 bytes
INFO:HST: [Host] Closing server socket
INFO:HST: chip=0 cpu=0 halted, status=0x0

```

Now we redo the block size to encourage the neural network to fit on two chips. As you will see however, it will error instead

```

In [8]:
block_sizes = [None, (8, 8, 4), (4, 4, 16), None]
conv_layers = [layer for layer in ann_model.layers if "conv" in str(layer).lower()]
for layer, block_size, layer_idx in zip(conv_layers, block_sizes, range(len(conv_layers))):
    if block_size == None: #
        continue
    output_shape = tuple(layer.output.shape[1:])
    with nengo_converter.net as net:
        net.config[net.ensembles[layer_idx+1]].block_shape = nengo_loihi.BlockShape(bloc

```

```

In [9]:
nengo_loihi.hardware.interface.HostSnip.recv_timeout = 1.0 # Default is 0.01 (10 ms)
nengo_loihi.hardware.interface.HostSnip.recv_retries = 100 # Default is 10
with nengo_loihi.Simulator(net, remove_passthrough=False, precompute = False, target="l
    print(f"Precompute = {loihi_sim.precompute}")
    # iterate over the timesteps

```

```
for i in range(10): # iterate over the batches
    loihi_sim.clear_probes() # prevents the buffer from filling up and crashing our
    loihi_sim.run(.25)
```

```
/homes/mjurado3/miniconda3/envs/loihi_vishal/lib/python3.9/site-packages/nengo_loihi/sim
ulator.py:159: UserWarning: Model is precomputable. Setting precompute=False may slow ex
ecution.
```

```
warnings.warn(
```

```
76
```

```
73
```

```
INFO:DRV: SLURM is being run in background
```

```
INFO:DRV: Connecting to 10.212.98.110:40359
```

```
INFO:DRV: Host server up.....Done 0.59s
```

```
INFO:DRV: Encoding axons/synapses.....Done 12.52s
```

```
INFO:DRV: Compiling Embedded snips....Done 1.30s
```

```
INFO:DRV: Compiling Host snips.....Done 1.13s
```

```
INFO:DRV: Compiling MPDS Registers....Done 1.35ms
```

```
INFO:HST: Args chip=0 cpu=0 /homes/mjurado3/miniconda3/envs/loihi_vishal/lib/python3.9/
site-packages/nxsdk/driver/compilers/../../../../temp/1641844238.992409/launcher_chip0_lmt
0.bin --chips=2 --remote-relay=1
```

```
INFO:HST: Args chip=1 cpu=0 /homes/mjurado3/miniconda3/envs/loihi_vishal/lib/python3.9/
site-packages/nxsdk/driver/compilers/../../../../temp/1641844238.992409/launcher_chip1_lmt
0.bin --chips=2 --remote-relay=1
```

```
INFO:HST: Nx...
```

```
INFO:DRV: Booting up.....Done 1.16s
```

```
INFO:DRV: Encoding probes.....Done 1.51ms
```

```
Precompute = False
```

```
INFO:DRV: Transferring probes.....Done 6.13ms
```

```
INFO:DRV: Configuring registers.....Done 2.46s
```

```
INFO:DRV: Transferring spikes.....Done 0.80ms
```

```
INFO:HST: [Host] Listening for client
```

```
INFO:HST: [Host] Connected to client
```

```
INFO:HST: [Host] Chip not responding. Ending simulation
```

```
INFO:HST: [Host] Wrote superhost shutdown signal: 8192 bytes
```

```
INFO:HST: [Host] Closing server socket
```

```

-----
RuntimeError                                Traceback (most recent call last)
/tmp/ipykernel_848125/1965710129.py in <module>
      7     for i in range(10): # iterate over the batches
      8         loihi_sim.clear_probes() # prevents the buffer from filling up and crash
ing our run
----> 9         loihi_sim.run(.25)
      10
      11

~/miniconda3/envs/loihi_vishal/lib/python3.9/site-packages/nengo_loihi/simulator.py in r
un(self, time_in_seconds)
      347         steps,
      348     )
--> 349         self.run_steps(steps)
      350
      351     def run_steps(self, steps):

~/miniconda3/envs/loihi_vishal/lib/python3.9/site-packages/nengo_loihi/simulator.py in r
un_steps(self, steps)
      360         raise SimulatorClosed("Simulator cannot run because it is closed.")
      361
--> 362         self._runner.run_steps(steps)
      363         self._n_steps += steps
      364         logger.info("Finished running for %d steps", steps)

~/miniconda3/envs/loihi_vishal/lib/python3.9/site-packages/nengo_loihi/simulator.py in l
oihi_bidirectional_with_host(self, steps)
      590         self.host.step()
      591         self._host2chip(self.loihi)
--> 592         self._chip2host(self.loihi)
      593         self.timers.stop("run")
      594

~/miniconda3/envs/loihi_vishal/lib/python3.9/site-packages/nengo_loihi/simulator.py in _
chip2host(self, sim)
      428
      429     def _chip2host(self, sim):
--> 430         sim.chip2host(self.probes_receivers)
      431
      432     @staticmethod

~/miniconda3/envs/loihi_vishal/lib/python3.9/site-packages/nengo_loihi/hardware/interfac
e.py in chip2host(self, probes_receivers)
      376         assert self.host_snip.connected
      377
--> 378         raw_data = self.host_snip.recv_bytes(self.bytes_per_step)
      379
      380         # create views into data for different chips

~/miniconda3/envs/loihi_vishal/lib/python3.9/site-packages/nengo_loihi/hardware/interfac
e.py in recv_bytes(self, bytes_expected)
      783         logger.debug("Received %d bytes, last_val=%d", len(data), last_val)
      784         if last_val == -1:
--> 785             raise RuntimeError("Received shutdown signal from chip")
      786
      787         return np.frombuffer(data, dtype=np.int32)

RuntimeError: Received shutdown signal from chip

```

In []: