Micron FPGA test on protoDUNE-SP

Manuel J. Rodriguez, Saul Alonso

R N K K N N

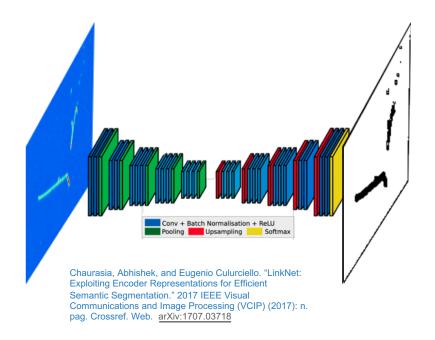
OUR PLANS

Data selection and trigger generation

- Focus on identifying areas of interest where there is activity on the detector.
- Fully Convolutional Networks to do image segmentation (**UNets**).
- Input: raw signals.

Micron

- **Goal**: checking the raw signals to get information from the waveforms.
 - Locate where there are hits!



OUR PLANS

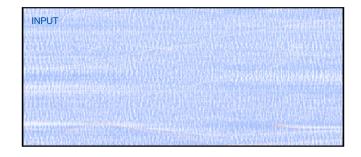
Data selection and trigger generation

- Focus on identifying areas of interest where there is activity on the detector.
- Fully Convolutional Networks to do image segmentation (**UNets**).
- Input: raw signals.

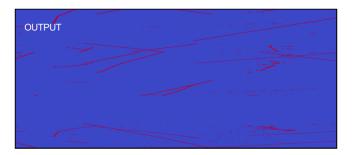
Micron

PLATFORM

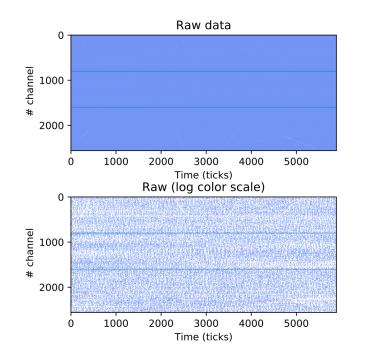
- **Goal**: checking the raw signals to get information from the waveforms.
 - Locate where there are hits!



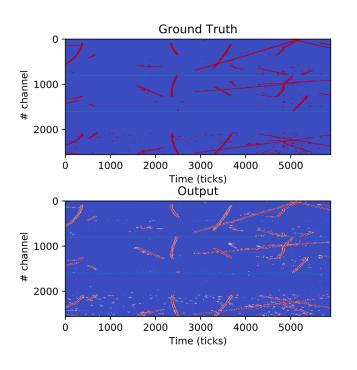




OFFLINE RESULTS



6/26/20

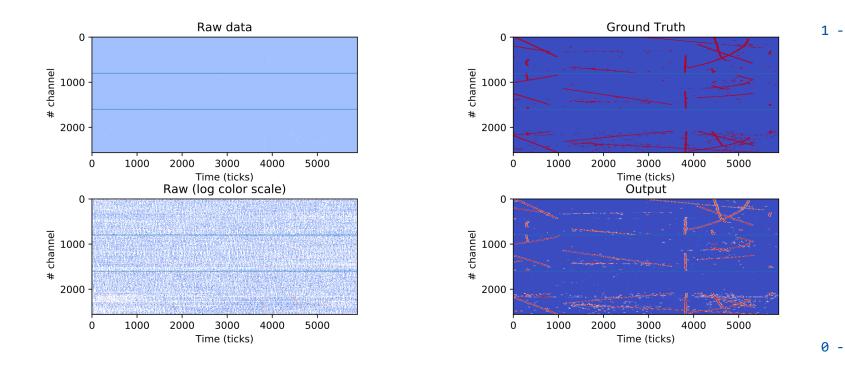


0 -

4

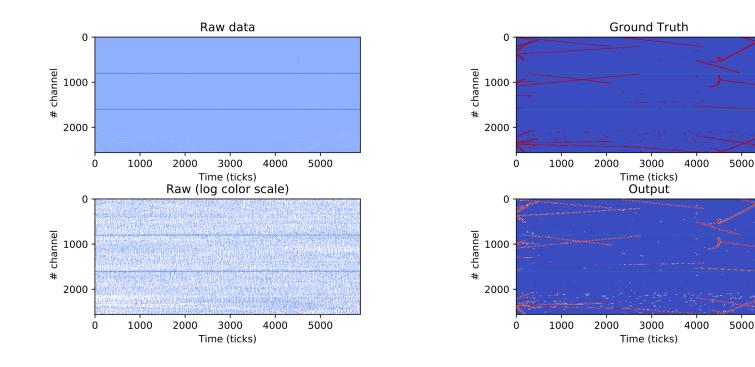
1 -

OFFLINE RESULTS



VELTERN CERN 6/26/20

OFFILINE RESULTS



0 -

1 -

ONLINE SOLUTIONS

- We aim to find these Regions of Interest (RoI) on real time.
- To analyze a whole trigger window of 3 ms we need to run the inference over 15'360'000 pixels (2560 channels times 6000ms clock ticks)
- After some research and reducing the network to its minimum, this cannot be done with all the incoming data. We need triggered data.
- We our goal is to run it at 12.5 Hz ,meaning that we have 80 ms to run the inference, per trigger window.

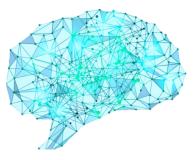
MICRON DLA

Direct deployment of neural networks on the inference engine

Micron Deep Learning Accelerator^[1]:

- No HDL programming.
- Natively supported neural networks.
- Most of the common layers are supported.
- Any framework that supports export to ONNX.
- Inference engine as an accelerator.

Micron



"Machine learning powers your world"

^[1]https://fwdnxt.com/

INFERENCE ENGINE

An FPGA ready for machine learning!

Micron Advanced Computing Solutions (ACS)

SB-852^[1]:

Micron

- Xilinx Virtex Ultrascale+ UV9P.
- 64GB DDR4 SODIMM (up to 512GB).
- 2GB Hybrid Memory Cube.
- 2 QSFP transceiver connectors.
- PCIe x16 Gen3 to the host.



• With the 2 Clusters version, the inference will take 700ms.

Х

INFERENCE ENGINE

An FPGA ready for machine learning!

Micron Advanced Computing Solutions (ACS)

AC-511 (x3)^[1]:

Micron

- Xilinx Virtex Ultrascale+ UV7P.
- 16GB DDR4 SODIMM
- 2GB Hybrid Memory Cube.
- PCIe x8 Gen3 to the host.
- SDAccel (OpenCL) support
- With the 4 Clusters version,
- the inference will take 100ms.







INFERENCE ENGINE

An FPGA ready for machine learning!

Micron Advanced Computing Solutions (ACS)

AC-511 (x3)^[1]:

Micron

- Xilinx Virtex Ultrascale+ UV7P.
- 16GB DDR4 SODIMM
- 2GB Hybrid Memory Cube.
- PCIe x8 Gen3 to the host.
- SDAccel (OpenCL) support
- With the 4 Clusters version,
- the inference will take 100ms.

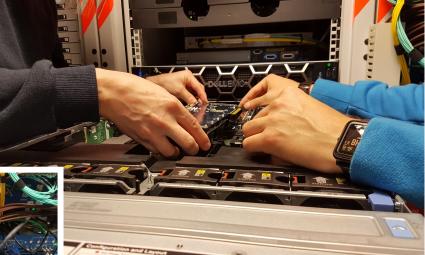
openlab



Almost nominal!

[1] https://www.micron.com/products/advanced-solutions/advanced-computing-solutions/ac-series-hpc-modules/ac-511









• We installed the driver for the Micron board and...

• We lost np04-srv-028

GNU GRUB version 1.99,5.11.0.175.1.0.0.13.18988

Minimal BASH-like line editing is supported. For the first word, TAB lists possible command completions. Anywhere else TAB lists possible device or file completions. ESC at any time exits.

grub> ls (hd0) (hd0,gpt9) (hd0,gpt2) (hd0,gpt1) (fd0)

grub> _



 We managed to fix the Grub, but Dracut wasn't happy either...

 The only solution:
 -> To call to our great System Administrators

	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
	dracut-initqueue[795]:						
316.3753581	dracut-initqueue[795]:	Warning:	dracut-initqueue	timeout	- starting	timeout	scripts
316.8918841	dracut-initqueue[795]:	Warning:	dracut-initqueue	timeout	- starting	timeout	scripts
317.4090091	dracut-initqueue[795]:	Warning:	dracut-initqueue	timeout	- starting	timeout	scripts
	dracut-initqueue[795]:						
318.4427791	dracut-initqueue[795]:	Warning:	dracut-initqueue	timeout	- starting	timeout	scripts
318.9595743	dracut-initqueue[795]:	Warning:	dracut-initqueue	timeout	- starting	timeout	scripts
318.959647]	dracut-initqueue[795]:	Warning:	Could not boot.				
Start	ing Setup Virtual Conso	le					
OK] Start	ed Setup Virtual Consol	е.					
	ing Dracut Emergency Sh						
karning: /dev/	disk/by-id/md-uuid-52a1	63a3:9270	0b33:b577e39f:a61	f14dc does	s not exis	t	
enerating "/r	un/initramfs/rdsosrepor	t.txt"					
ntening emeng	encu mode Fyit the she	11 to com	tinue				

Entering emergency mode. Exit the shell to continue. Type "journalctl" to view system logs. You might want to save "runrinitramfszrdsosreport.txt" to a USB stick or /boot after mounting them and attach it to a bug report.

acut:/# exit



- The diagnosis was that the driver module (which is compiled on the host to ensure compatibility) was corrupted. Therefore the system failed to load the module and all its dependencies.
- We reinstalled it and it worked.



TESTING THE BOARD ON srv-028

- We tested it over and over and over again.
- However, every time we tried to run the FPGA it was throwing a "bad fpga seq"
- At this point Micron joined the test.
- We tried together to debug it without any success. Even with a simple demo firmware on the FPGA it was failing.
- They thought that it could be a hardware failure.

Micron

[mjrodrig@np04-srv-028 ProtoDUNE-scripts]\$./threadedbatchdemo -i test/ -s tinylinknet_20200528.bin -r 1024x2560x1 -f 3 -C 4 -B
ie_init: Initialize Micron DLA system DLA binary to be read is tinylinknet_20200528.bin
Using FPGA 0x511 Device 0511 ^C

+0.000002] pico: couldn't send 'read' command to system PicoBus: -10011 +9.303503] pico: interrupted while waiting for dma

TESTING THE BOARD ON srv-028

- We tried replacing one of the three FPGA, that seems faulty, but we still were having the same issue.
- Micron is still investigating this issue.
- Solution: Try the old SB-852
 Not ideal at all.

[mjrodrig@np04-srv-028 ProtoDUNE-scripts]\$./threadedbatchdemo -i tinylinknet_20200528.bin -r 1024x2560x1 -f 3 -C 4 -B	i test/ -s	
ie_init: Initialize Micron DLA system DLA binary to be read is tinylinknet_20200528.bin		
Using FPGA 0x511 Device 0511 ^C		

+0.000002] pico: couldn't send 'read' command to system PicoBus: -10011 +9.303503] pico: interrupted while waiting for dma

TESTING THE BOARD ON srv-028

- We didn't manage to use the 4 Cluster version on the SB-852 (actually this firmware was experimental)
- With the 2 Cluster it takes 700 ms per trigger window
 Trigger rate at 1.4 Hz

)28 ProtoDUNE-scripts]\$./threadedbatchdemo -i test/ -s 3.bin -r 1024x2560x1 -f 3 -C 4 -B
ie_init: Initialize	Micron DLA system d is tinylinknet_20200528.bin
Using FPGA 0x511 Dev ^C	rice 0511
[]un16 16:31] nico:	bad fpga seq for fpga 1 stream 254! expected 0x190,

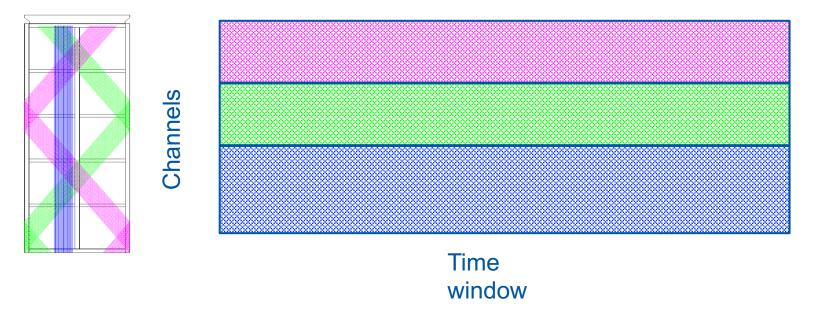
- got 0x180. last_host_seq: 0x180 (desc seq: 0x190)
 [+0.000003] pico:pico_newfw(): pico_newfw_internal() return error: 10011
- +0.000002] pico: couldn't send 'read' command to system PicoBus: -10011 +9.303503] pico: interrupted while waiting for dma

• After all the issues, we managed to send data to the FPGA, in a one shot approach.

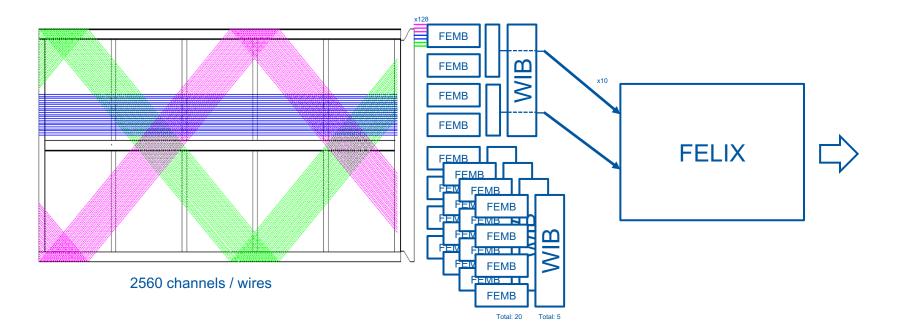
• However, we found a totally different issue.



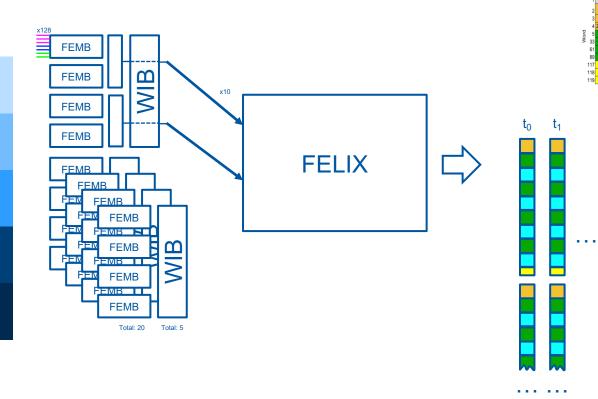
• Our images in our dataset are like this:











PLATFORM PLATFORM



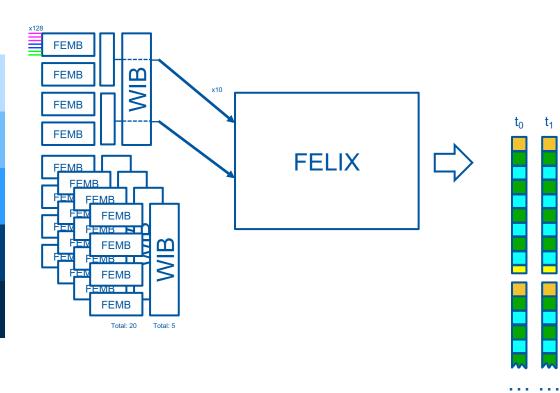
 t_{6000}

. . .

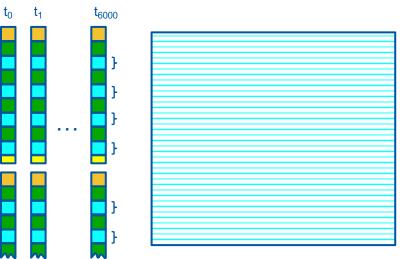


6/26/20

6/26/20



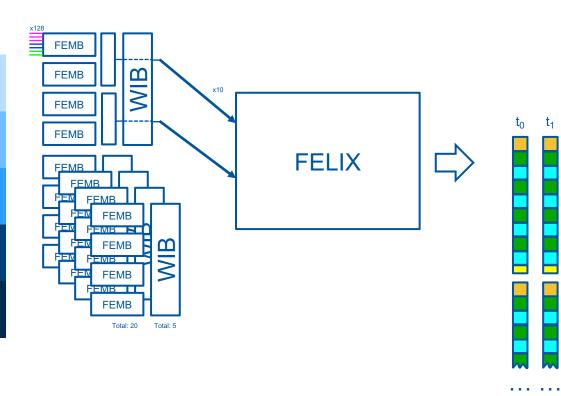
				B	lit												
		~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	BIL														
	0	0x0 0x0 0x0 SOF												2424 42424 40 0 0		くららきうなへの Stream 2 Stream 1	
	1	Reserved (8)	SlotNo		FiberNoVersion = 0x1	0x0	Ħ		1	ChkS	im B [7:0] ChkSm A [7:0]		Reserved (8)		ERR 2	ERR	
	1								2	2 COLDDATA Co			nvert Count		ChkSm B [15:8]		A [15:8]
	2	WIB	Errors		Reserved (14)				3	3 Reserved (16)						Register	
	3			Timestar	mp (31:0)	[31:0]			- 4	HDR8	HDR6	HDR7	HDR5	HDR4	HDR2	HDR3	HDR1
	4	Timestamp [62:	48] or W	1B counter	Timestan	11	+-		p 5	ADC2 CH2[3:0	ADC2 ADC2 ADC1 ADC1 CH2[3:0] CH1[11:8] CH2[3:0] CH1[11:			ADC2 CH1[7:0]		ADC1 CH1[7:0]	
Word	5		COLDATA Block 1							ADC2	CH3[7:0]	ADC1	CH3[7:0]	ADC2 CH2[11:4]		ADC1 C	H2[11:4]
3	33			· · ·						ADC2	ADC2	ADC1	ADC1				
	61		COLDATA Block 3							ADC2	CH4[11:4]	ADC1 CH4[11:4]		CH4[3:0] CH3[11:8]		CH4[3:0]	
	89																
1	17	0x0 CRC-20 [19:0] E															
1	18	0x00		0x00	0x00	K28.5			28	ADC8	CH8[11:4]	ADC7 C	H8[11:4]	ADC8 CH8(3:01	ADC8 CH7[11:8]	ADC7 CH8[3:0]	ADC7 CH7[11:8]
1	19	0x00		0x00	0x00	K28.5								0.00	0(11.0)	0.00	011111101



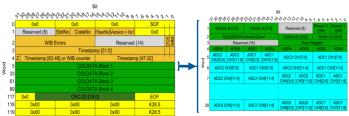
. . .

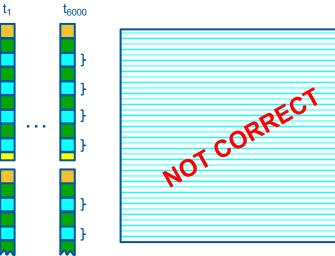
#### ADC values for the time window

6/26/20



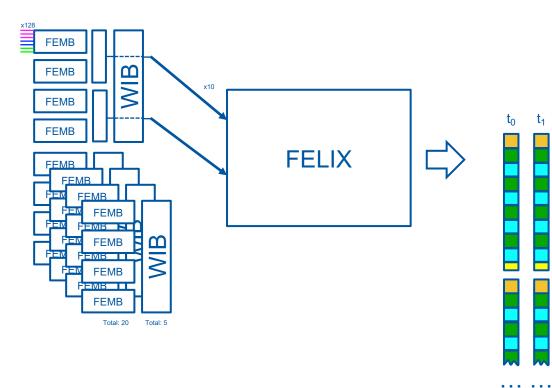
PLATFORM PLATFORM





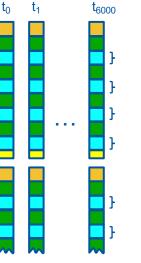
. . .

#### ADC values for the time window

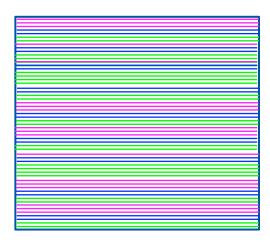


6/26/20

			6	Bit												
		1000000 march	᠂ᡔ᠋᠕᠕᠅ᠬᡐ᠙ᡐᡧᡧ᠙	Bit												
					\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$											
	0	0x0	0x0	0x0	SOF		<u> </u>		ChkSm B [7:0]		ChkSm A [7:0]		Reserved (8)		Stream 2	Stream 1
	1	Reserved (8)	SlotNo CrateNo	FiberNo Version = 0x1	0x0										ERR	ERR
			-		8			2	0		A Convert Count		ChkSm B [15:8]		ChkSm legister	A [15:8]
	2	WIB	Errors	Reserved (14)				3		Reserv		_				
	3		Timesta	mp [31:0]				4	HDR8	HDR6	HDR7	HDR5	HDR4	HDR2	HDR3	HDR1
	4 2	Z Timestamp [62:	48] or WIB counter	np [47:32]		+++		p 5	ADC2 CH2[3:0]	ADC2 CH1[11:8]	ADC1 CH2[3:0]	ADC1 CH1[11:8]	ADC2	CH1[7:0]	ADC1 C	CH1[7:0]
Word	5		COLDATA Block 1							CH3[7:0]	ADC1 CH3[7:0]		ADC2 CH2[11:4]		ADC1 C	H2[11:4]
3	33			· · ·						ADC2	ADC2	ADC1	ADC1			
	61	COLDATA Block 3							ADC2	CH4(11:4)	ADC1 CH4[11:4]		CH4[3:0] CH3[11:8			
	89		COLDAT													
1	117	0x0 CRC-20 [19:0] EOF														
1	118	0x00	0x00	0x00	K28.5			28	ADC8	CH8[11:4]	ADC7 C	H8(11:4)	ADC8 CH8(3:01	ADC8 CH7[11:8]	ADC7 CH8[3:0]	ADC7 CH7[11:8]
1	119	0x00	0x00	0x00	K28.5								0.000	[0111]11.0]	0.000	0.0110.001

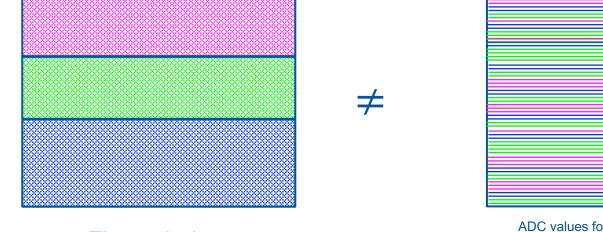


. . .



ADC values for the time window

Channels



Time window

ADC values for the time window



- Removing the headers is fine
- Reordering the data (2560 channels time 6000 ticks) is not.
- Possible solutions:
  - Retrain the network using the online channel number
  - Do the reorder on FPGA (FELIX or Inference Engine)

# CONCLUSIONS

- We wanted to test the integration of the Micron DLA on the protoDUNE DAQ chain.
- The hardware we used was an unreleased version made for this test with some issues unseen before. Thanks to the test, Micron can study it and debug it to make the system more robust.
- We faced as well a different issue not taken into account. In my opinion, it was great that we worked with online raw data. This gave us a much better understanding on how that data is coming from the detector.
- It's a pity that we don't have more time to test. However, thanks to the binary data recorded we can continue evolving the system.

### **THANK YOU**



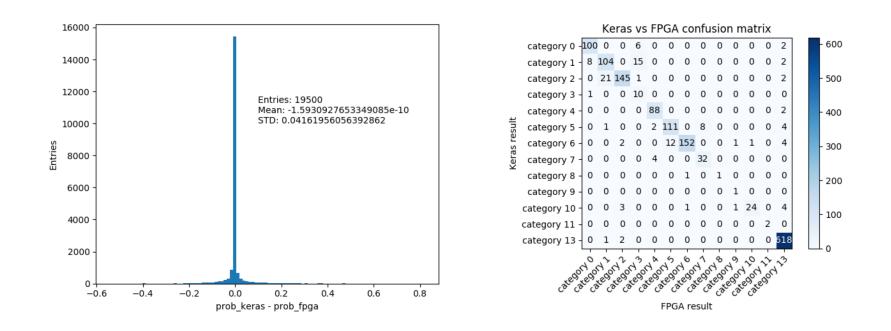




#### **GPU-FPGA RESULTS COMPARISON**

#### How good our FPGA behaves

Neutrino





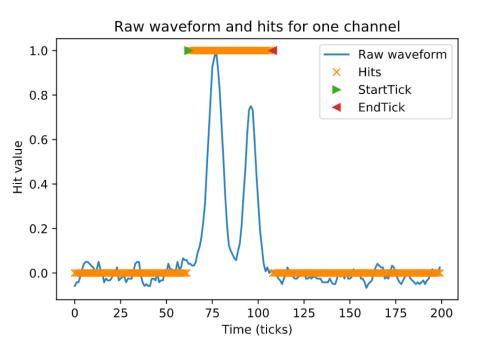
# **Our dataset**

• On the hits file we have:

(int)hit.Channel(), hit.StartTick(), hit.EndTick(), (int)hit.SummedADC(), (int)hit.RMS()

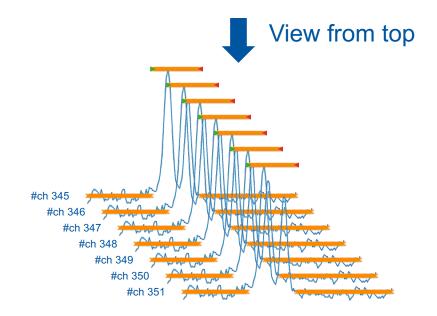
 We take the StartTick and EndTick and we mark the whole range as

hit(channel,[startTick,endTick]) = TRUE



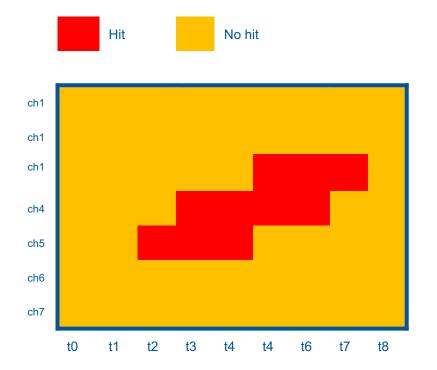
# **Region of interest**

- Once we have the hits for all the channels.
- We artificially augment the hits area in time and channels to get our region of interest.
- ∀ *i*, *j*: If hit(i,j) == 1
  - hit(i +1, j) = 1
  - hit(i -1, j) = 1
  - hit(i, j+1) = 1
  - hit(i, j-1) = 1
  - hit(i +1, j+1) = 1
  - [...]



# **Region of interest**

- Once we have the hits for all the channels
- We artificially augment the hits area in time and channels to get our region of interest
- ∀ *i*, *j*: If hit(i,j) == 1
  - hit(i +1, j) = 1
  - hit(i -1, j) = 1
  - hit(i, j+1) = 1
  - hit(i, j-1) = 1
  - hit(i +1, j+1) = 1
  - [...]



# **Region of interest**

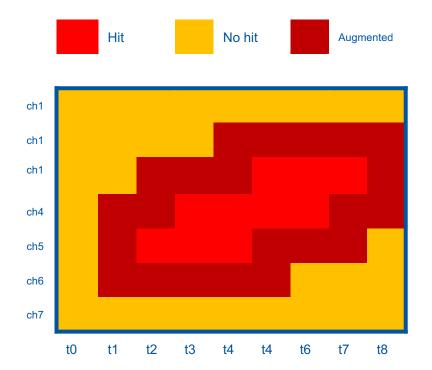
- Once we have the hits for all the channels
- We artificially augment the hits area in time and channels to get our region of interest
- ∀ *i*, *j*: If hit(i,j) == 1
  - hit(i +1, j) = 1
  - hit(i -1, j) = 1
  - hit(i, j+1) = 1
  - hit(i, j-1) = 1
  - hit(i +1, j+1) = 1
  - [...]

Micron

Neutrino PLATFORM

• We use the augmented area as our ground truth for the neural network

6/26/20



35