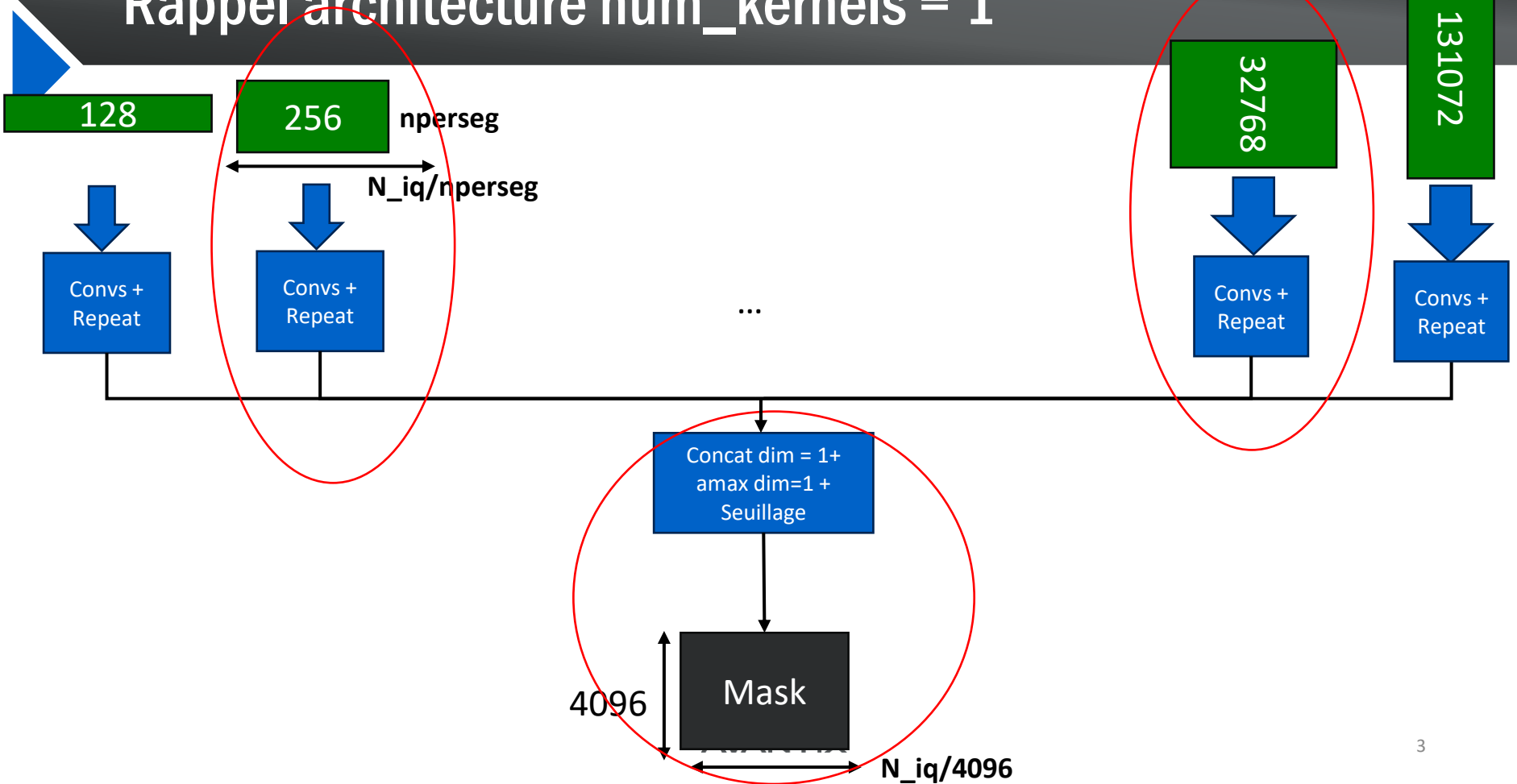
AVANTIX

Avancement Simulateur IQ

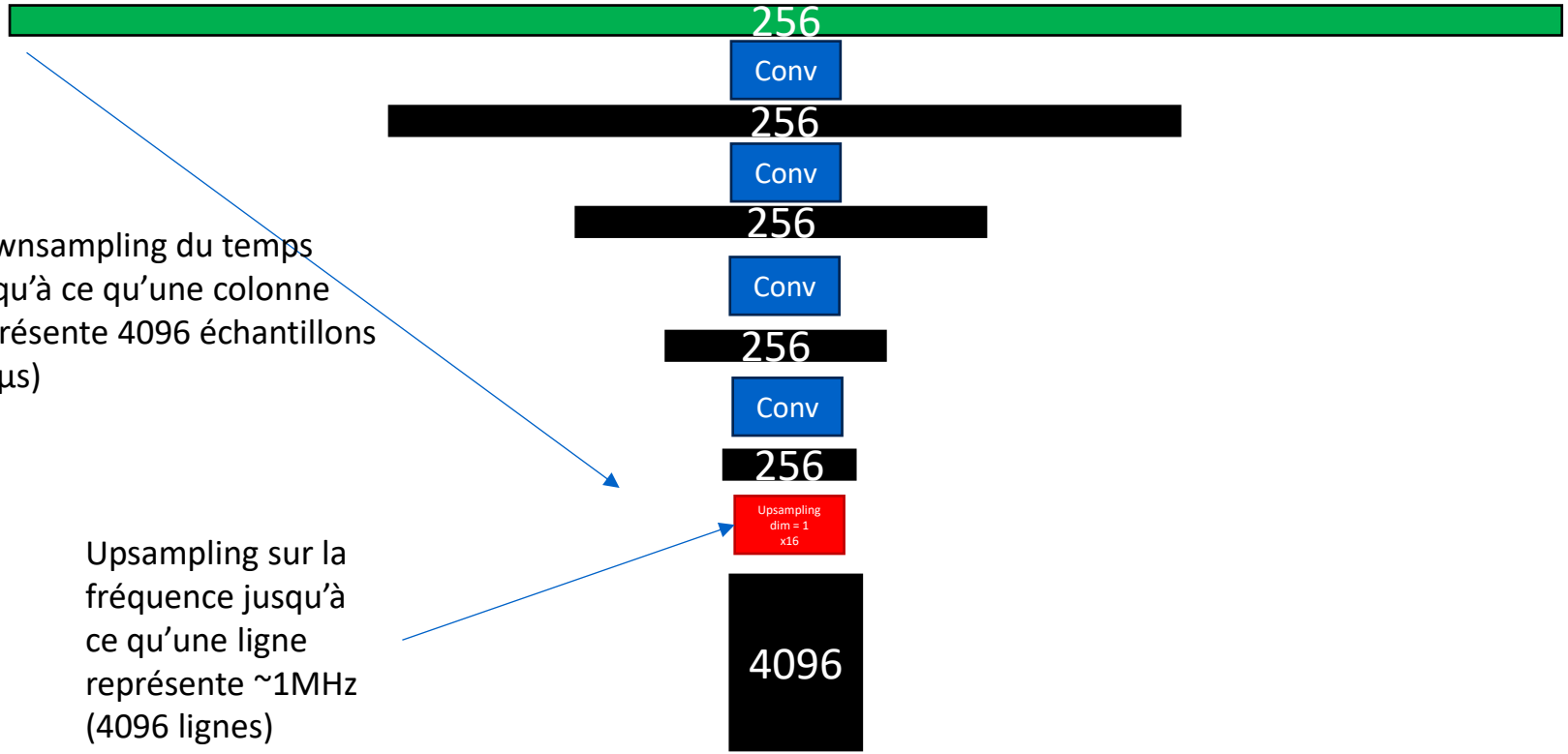
AVANTIX

Revue architecture

Rappel architecture num_kernels = 1



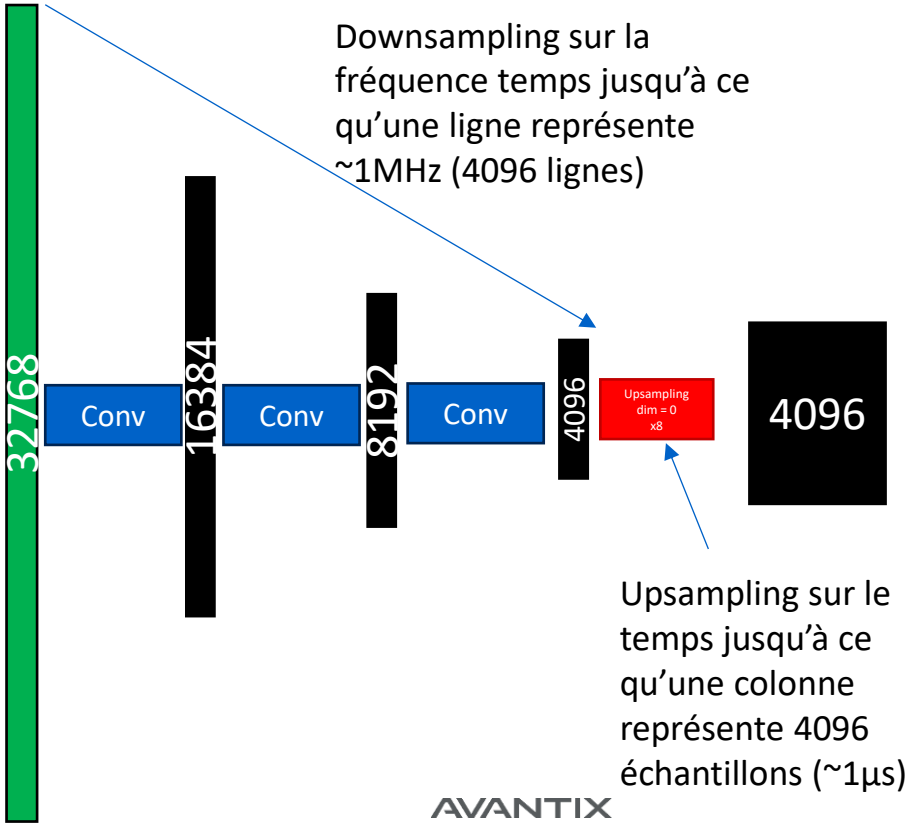
Rappel architecture modèle nperseg < 4096



Downsampling du temps jusqu'à ce qu'une colonne représente 4096 échantillons (~1µs)

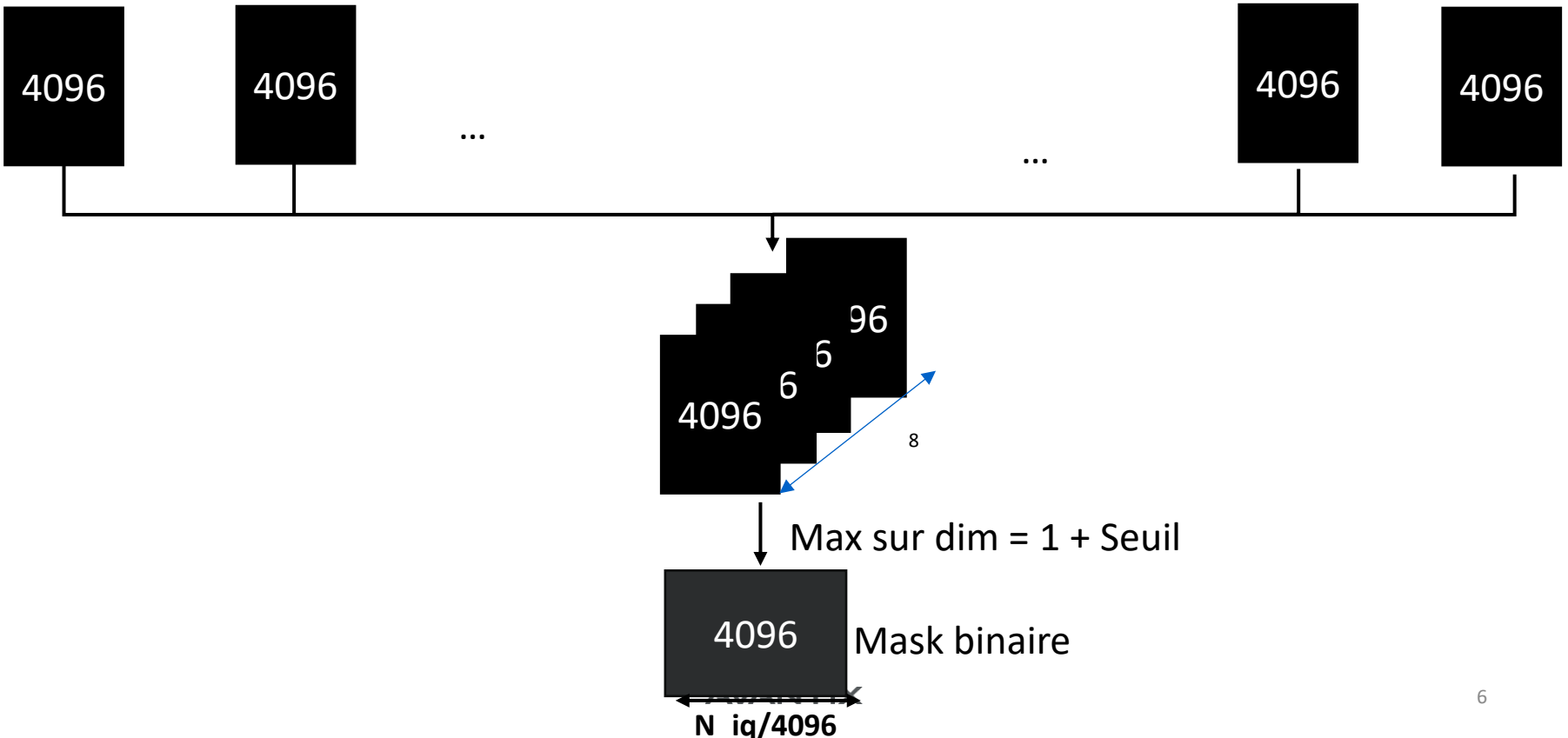
Upsampling sur la fréquence jusqu'à ce qu'une ligne représente ~1MHz (4096 lignes)

Rappel architecture modèle nperseg > 4096



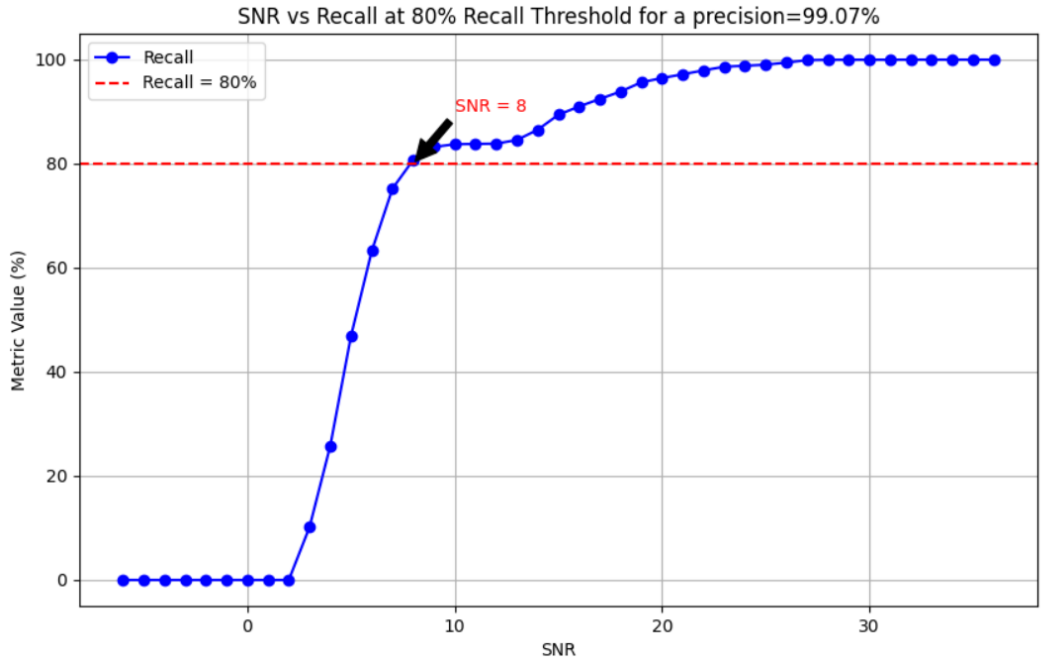
Rappel architecture

128 256 32768 131072



Objectif : Améliorer recall

run_2025-01-24_12-15-02



Voies pour améliorer le recall :

1. **Trouver des architectures plus adaptées au problème**
2. **Capper le SNR eff à 17dB**
3. **BT avec moins d'impulsions**
4. **Modèle plus gros avec FP8, TensorRT, et pruning**

Rappel architecture actuelle

NPERSEG	Détails branche
64	6 convs ayant 1 kernel chacun
128	5 convs ayant 1 kernel chacun
256	4 convs ayant 1 kernel chacun
512	3 convs ayant 1 kernel chacun
1024	2 convs ayant 1 kernel chacun
4096	1 conv ayant 1 kernel chacun
16384	2 convs ayant 1 kernel chacun
131072	5 convs ayant 1 kernel chacun

Modèle à 199 paramètres
RTR V100 = 0,95
RTR H100 = 0,56

Architecture 1 : Equilibrage du nombre de kernels par branche

NPERSEG	Détails branche
64	6 convs ayant 1 kernel chacun
128	5 convs ayant 1 kernel chacun et 2 pour le premier
256	4 convs ayant 1 kernel chacun et 3 pour le premier
512	3 convs ayant 1 kernel chacun et 4 pour le premier
1024	2 convs ayant 1 kernel pour le deuxième et 5 pour le premier
4096	1 conv ayant 6 kernel + 1 conv k=1, s=1, p=0
16384	2 convs ayant 1 kernel pour le deuxième et 5 pour le premier
131072	5 convs ayant 1 kernel chacun et 2 pour le premier

Modèle à 451 paramètres
RTR V100 = 3,45
RTR H100 = 1,45

Architecture 1bis (JDB)

NPERSEG	Détails branche
64	6 convs ayant 1 kernel chacun
128	5 convs ayant 1 kernel chacun
256	4 convs ayant 1 kernel chacun
512	3 convs ayant 1 kernel chacun
1024	2 convs ayant 1 kernel pour le deuxième et 4 pour le premier
4096	1 conv ayant 4 kernel + 1 conv k=1, s=1, p=0
16384	2 convs ayant 1 kernel pour le deuxième et 2 pour le premier
131072	5 convs ayant 1 kernel chacun et 2 pour le premier

Modèle à 299 paramètres
RTR V100 = 2,40
RTR H100 = 1,00

Architecture 2 : Uniformisation des branches

NPERSEG	Détails branche
64	1 conv (k=3, p=1, s=(2, 1)) + amax (dim=1) + MaxPool2D + UpSampling
128	1 conv (k=3, p=1, s=(2, 1)) + amax (dim=1) + MaxPool2D + UpSampling
256	1 conv (k=3, p=1, s=(2, 1)) + amax (dim=1) + MaxPool2D + UpSampling
512	1 conv (k=3, p=1, s=(2, 1)) + amax (dim=1) + MaxPool2D + UpSampling
1024	1 conv (k=3, p=1, s=(2, 1)) + amax (dim=1) + MaxPool2D + UpSampling
4096	1 conv (k=3, p=1, s=(1, 1)) + amax (dim=1) + Identity + Identity
16384	1 conv (k=3, p=1, s=(1, 2)) + amax (dim=1) + MaxPool2D + UpSampling
131072	1 conv (k=3, p=1, s=(1, 2)) + amax (dim=1) + MaxPool2D + UpSampling

Num kernels	Num parameters	RTR V100	RTR H100
1	**80**	**0,49**	
2	**160**	**0,87**	
3	240	2,29	
4	**320**	**1,58**	
5	400	3,75	
6	480	4,47	
8	**640**	**3,54**	
16	**1280**	**3,97**	
32	**2560**	**24,03**	**5,72**
64	**5120**	**191,25**	**39,97**

Architecture 3 : Resize + 1 conv2d + 1 conv2d chan reducer

Resize de tous les SXX à max(height) et max(width)
1 seule branche de conv

Modèle à 148 paramètres

*n_iq = 1 * 131072 * 32*

```
RuntimeError: input tensor has spatial dimension larger than the kernel capacity
```

ptrblck Nov 2020

This is a limitation in the CUDA launch config for the current algorithm. As a workaround you could use the CPU for these large shapes and create a feature request on GitHub.

Architecture 4 : Imbalanced

64 et 128 : 1 seul kernel
Pour tout le reste, plus de kernels et égaux entre les npersegs restants

NPERSEG	Détails branche
64	1 conv (k=3, p=1, s=(2, 1)) + amax (dim=1) + MaxPool2D + UpSampling
128	1 conv (k=3, p=1, s=(2, 1)) + amax (dim=1) + MaxPool2D + UpSampling
256	4 convs ayant 1 kernel chacun et 2 pour le premier
512	3 convs ayant 1 kernel chacun et 3 pour le premier
1024	2 convs ayant 1 kernel pour le deuxième et 4 pour le premier
4096	1 conv ayant 5 kernel + 1 conv k=1, s=1, p=0
16384	2 convs ayant 1 kernel pour le deuxième et 4 pour le premier
131072	5 convs ayant 1 kernel chacun

Modèle à 305 paramètres
RTR = 2,768

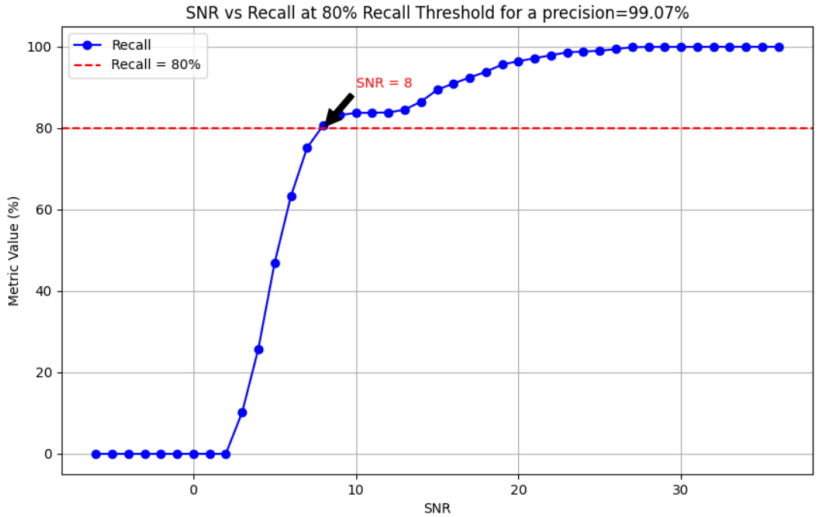
*n_iq = 4 * 131072 * 128*

AVANTIX

Cap 17dB sur SXX + Plus gros modèles

Objectif : Améliorer recall sur les bas SNR

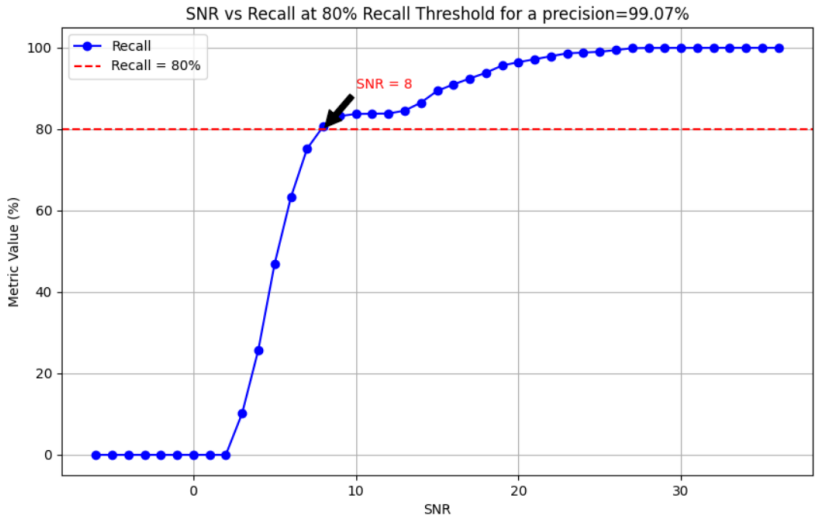
run_2025-01-24_12-15-02
uncapped & petit modèle (num_kernels=1)



run_2025-02-04_xx-xx-xx
capped & petit modèle (num_kernels=1)

Objectif : Améliorer recall sur les bas SNR

run_2025-01-24_12-15-02
uncapped & petit modèle (num_kernels=1)

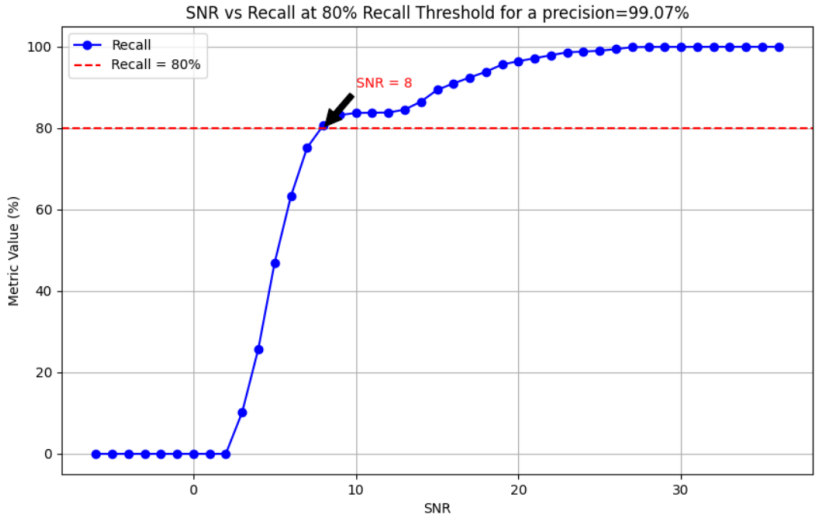


run_2025-02-04_xx-xx-xx uncapped
& gros modèle (num_kernels=4)

Objectif : Améliorer recall sur les bas SNR

run_2025-01-24_12-15-02
uncapped & petit modèle (num_kernels=1)

run_2025-02-04_xx-xx-xx
capped & gros modèle (num_kernels=4)



AVANTIX

MERCI