

Latence d'un lien Ethernet Gigabit : Mesure par TDC et DDMTD

Yassir MOUDDEN

- **IRFU** : Frédéric LOUIS, Hervé LE PROVOST, Frédéric CHATEAU, Bertrand VALLAGE, Shebli ANVAR, Eric ZONCA, Kevin MENAGER
- **NIKHEF** : Peter JANSWEIJER, Jan-Willem SCHMELLING, Sander MOS, Albert ZWART, Henk PEEK, Jelle HOGENBIRK
- **NESTOR** : Kostas MANOLOPOULOS
- **INFN** : Fabrizio AMELI



Immersion réussie d'un premier DOM (Digital Optical Module) le 17 avril 2013 à l'occasion de la remise à l'eau d'une ligne instrumentée d'Antares.

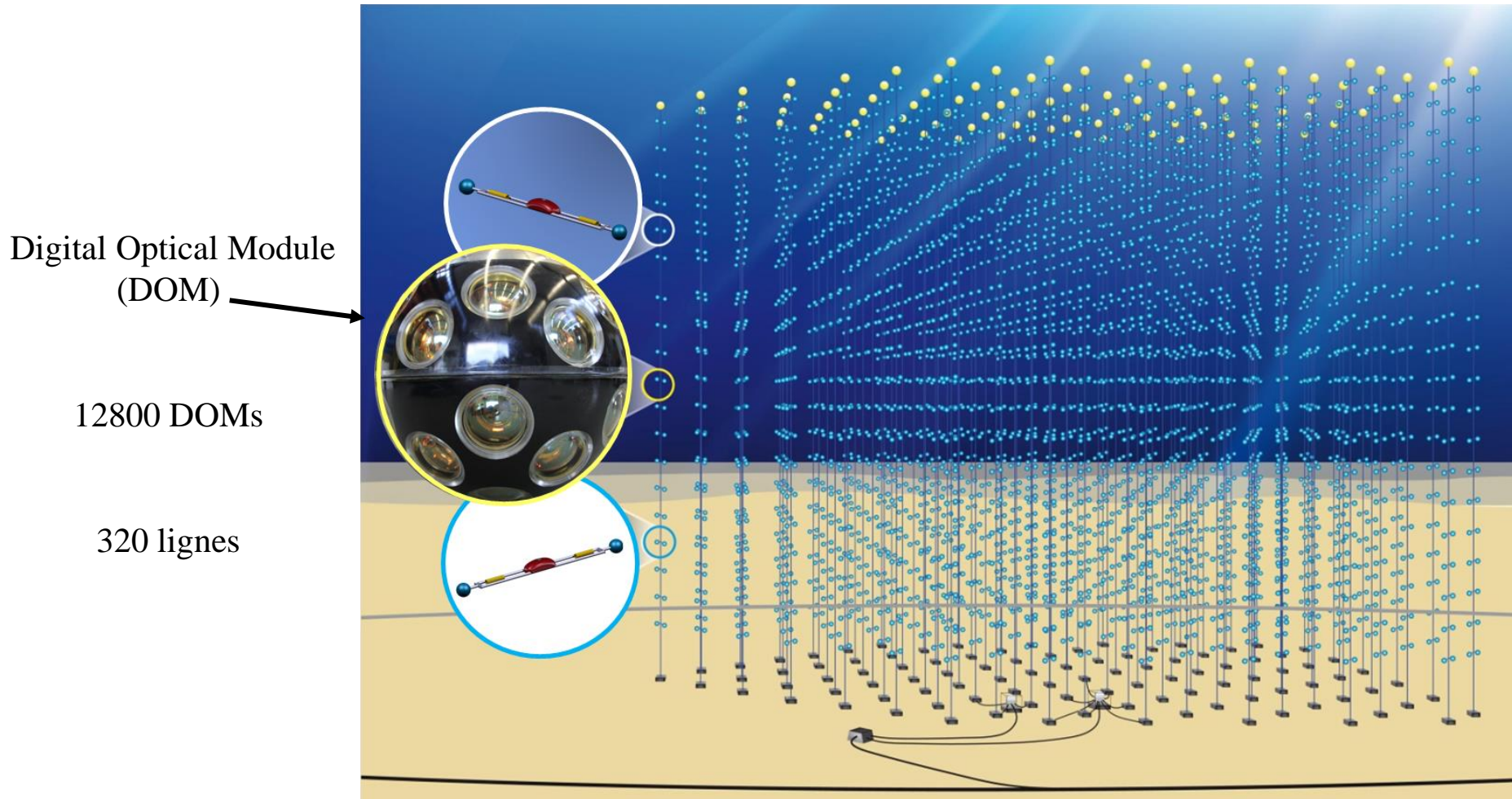
Trois autres DOMs immergés avec succès au large de la Sicile en Mai 2014.

Le projet de détecteur de neutrinos KM3NET

<http://www.km3net.org>

Technical Design Report (ISBN 978-90-6488-033-9)

Conceptual Design Report (ISBN 978-90-6488-031-5)



N.B. L'IRFU s'est retiré du projet fin 2012.

MOTIVATION : Distribution d'horloge et Synchronisation

“Syntonization”

Distribution d'une fréquence issue d'une source unique (GPS, horloge atomique) d'un maître vers un ou plusieurs esclaves distants.



Utiliser la fréquence *porteuse* des liens Gigabit codés 8b/10b.

(par exemple : SONET/SDH, Synchronous Ethernet, White Rabbit)

“Synchronization”

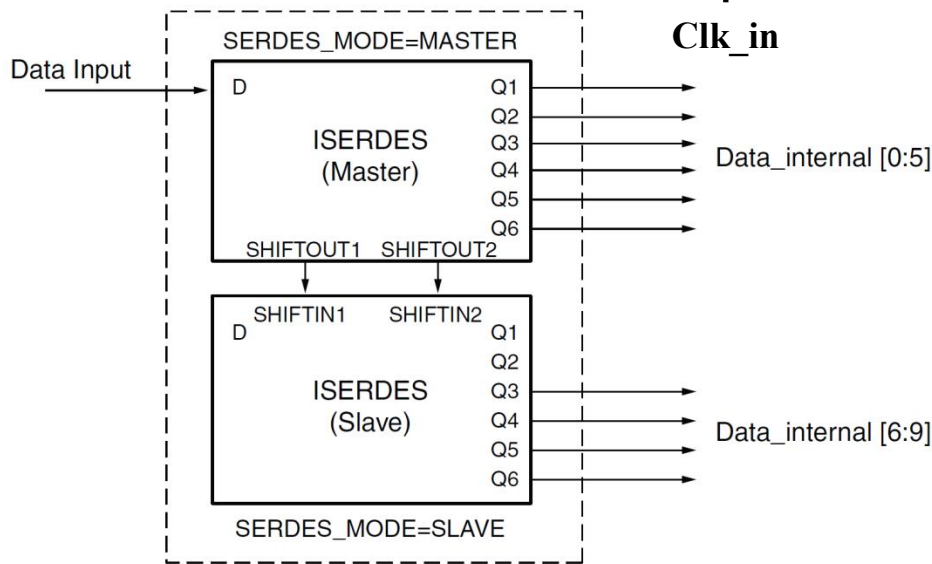
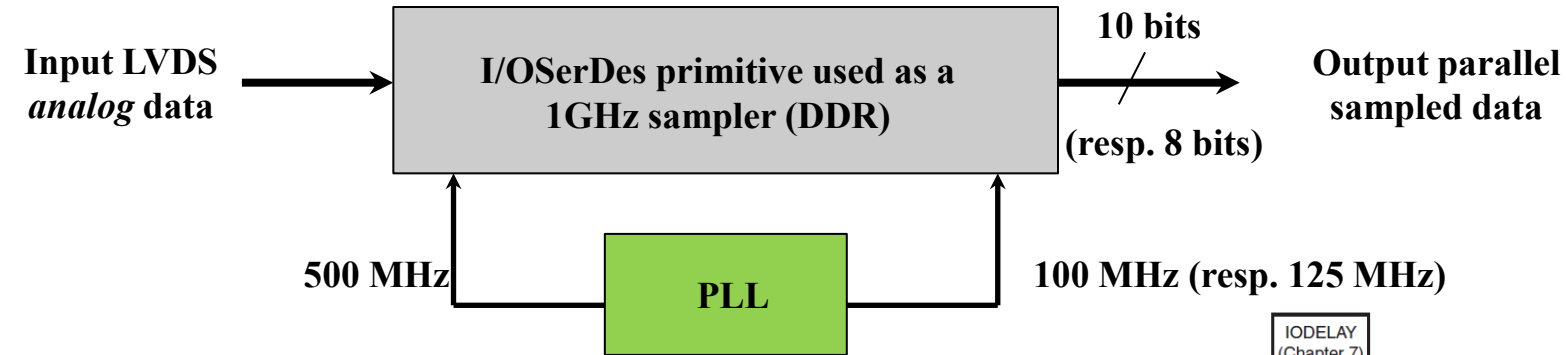
Distribution d'une phase / d'une date d'un maître vers un ou plusieurs esclaves avec une précision qui dépend des “*besoins*” de l'expérience (e.g. mieux que la nanoseconde) .



Besoin de commandes synchrones (*top départ*) dans la couche physique, sans modifier les couches supérieures du protocole Ethernet.

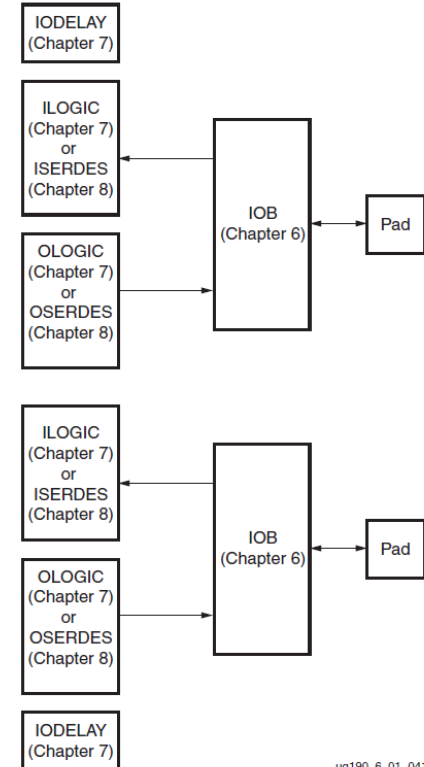
Besoin de pouvoir mesurer la latence A/R, répartir cette latence entre Aller et Retour, et ses variations.

Single channel 1 GHz sampler as TDC



ug361_c3_08_022309

Figure 3-9: Block Diagram of ISERDES Width Expansion

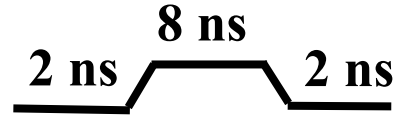


ug190_6_01_041106

Figure 6-1: Virtex-5 FPGA I/O Tile

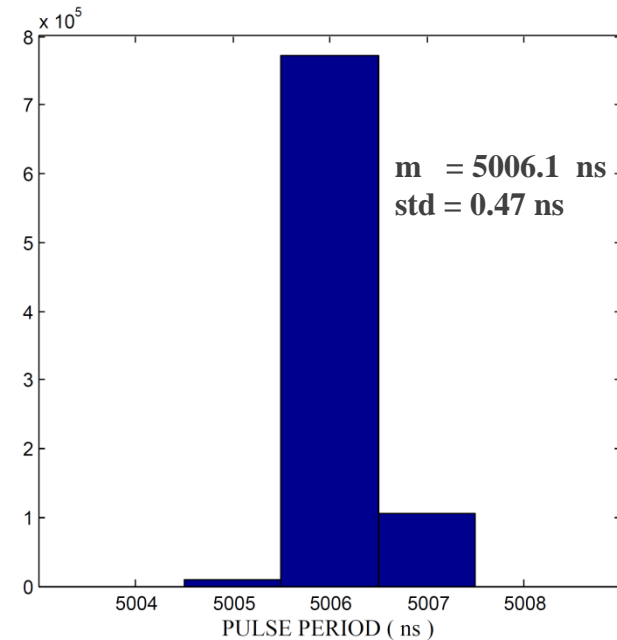
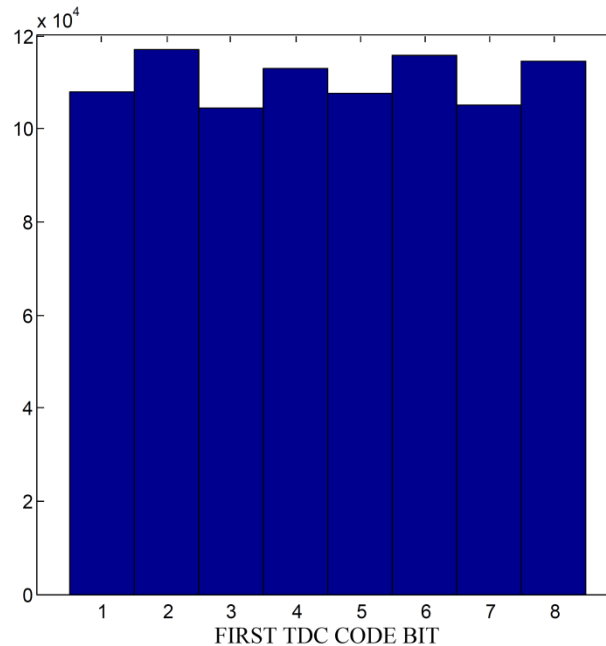
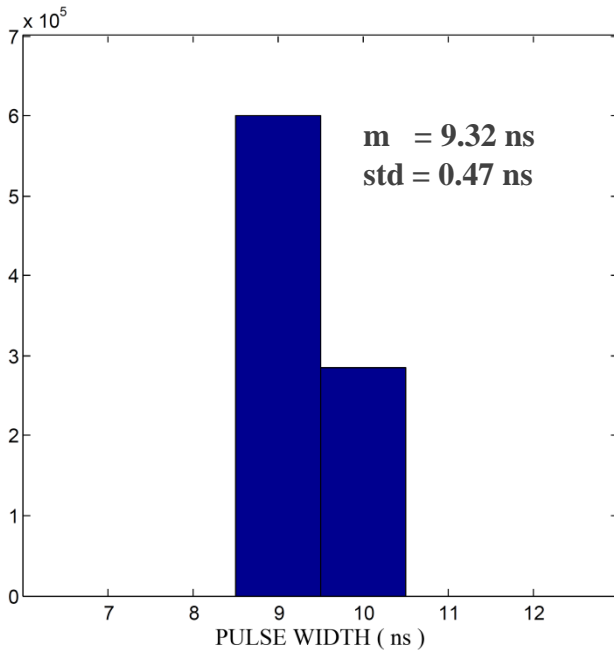
TDC : results (2)

Générateur d'impulsions :



Fréquence : 200 kHz

(LeCroy 64Xi 600 MHz donne std = 220 ps)

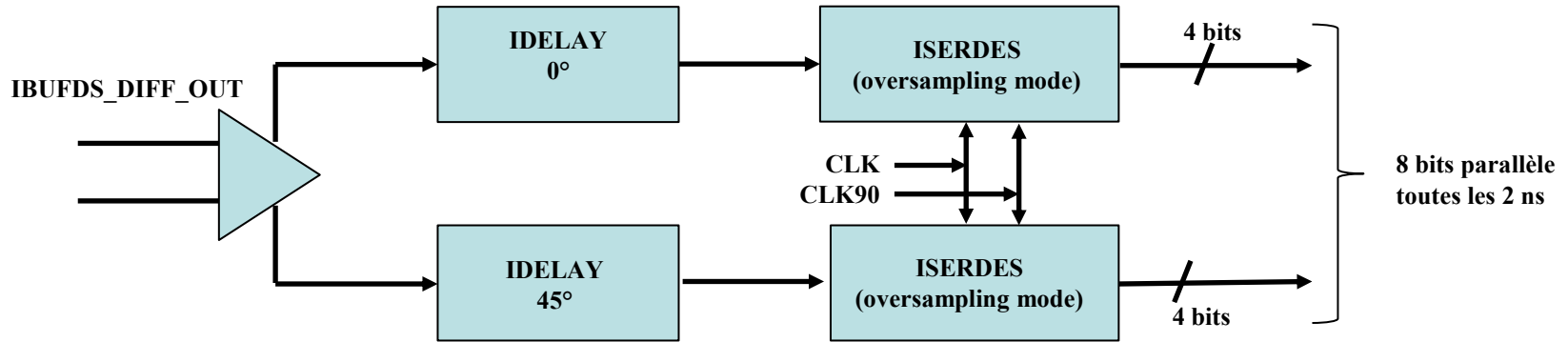


TDC 250 ps

- Plus que largement inspirée de

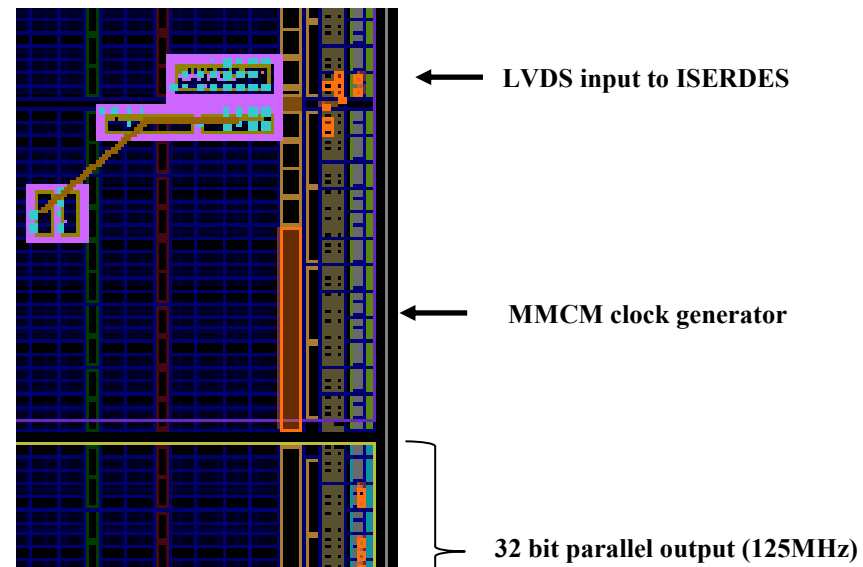
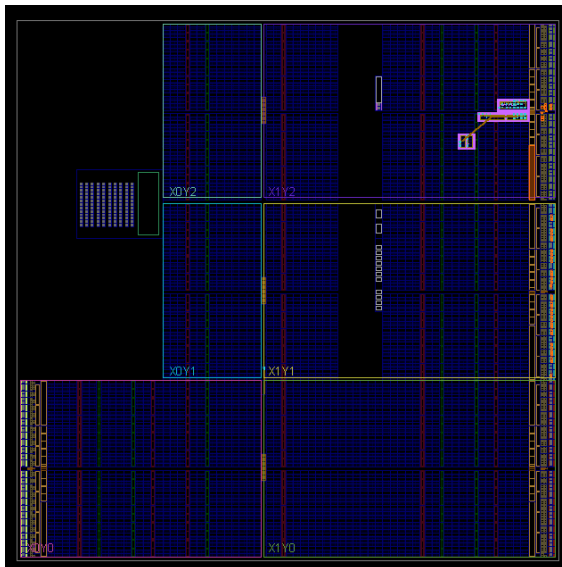


LVDS 4x Asynchronous Oversampling
Using 7 Series FPGAs
Author: Marc Defossez



CLK et CLK90 deux horloges à 500 MHz déphasées de 90°. Réalise un échantillonnage *régulier* toutes les 250 ps.

- Portage sur ZYNQ en cours



Digital Dual Mixing Time Difference (DDMTD)

- Mesurer un déphasage entre deux signaux periodiques de même fréquence F .
- Les signaux sont échantillonnés à une fréquence $F + e$.

➔ Dilatation du temps rend possible la mesure de déphasage par comptage.

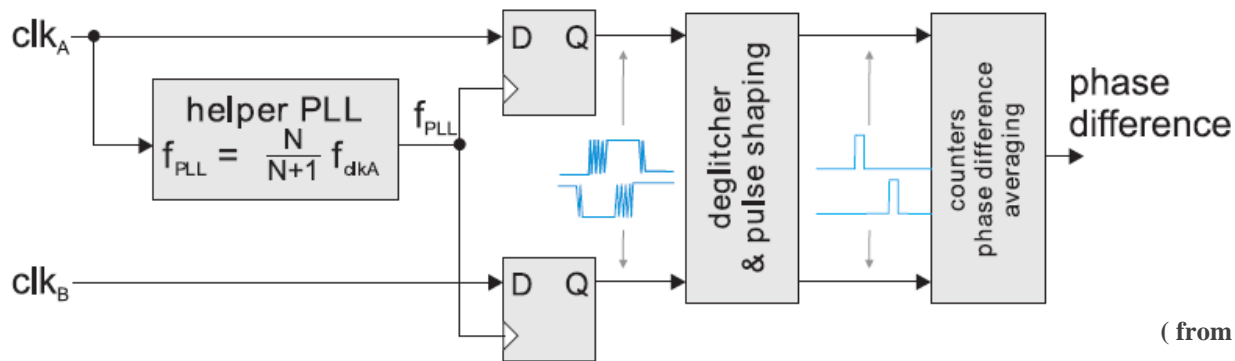
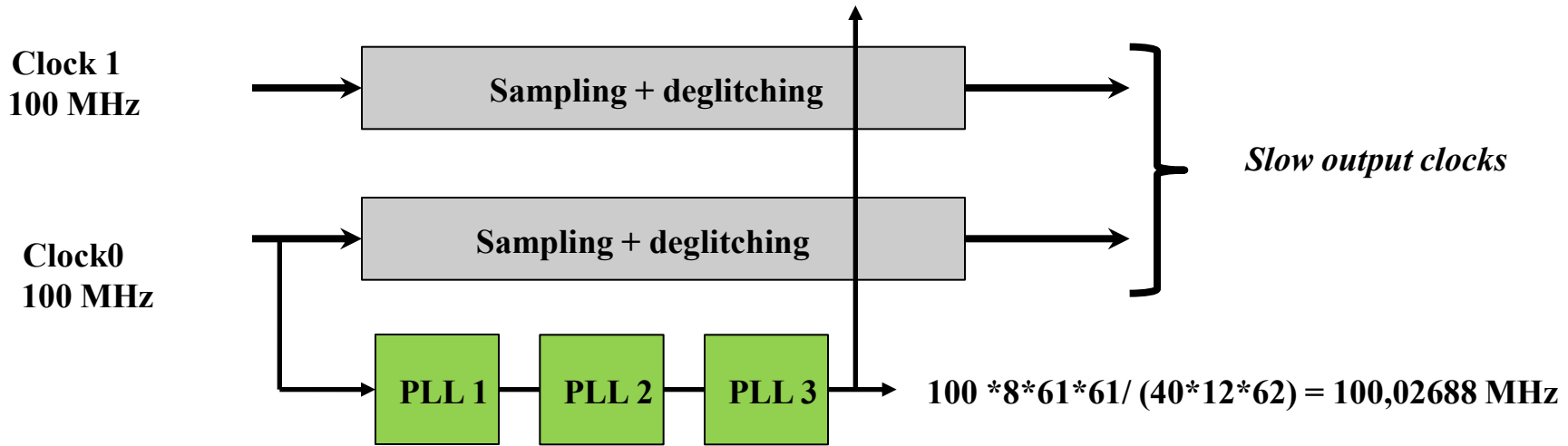


Figure 3.19. Structure of a digital DMTD phase detector

- PLL externe (e.g. White Rabbit) ou cascade de PLLs ?
- Compromis temps d'integration / precision de la mesure / proximité des fréquences.
- Meilleur filtrage des "glitches".

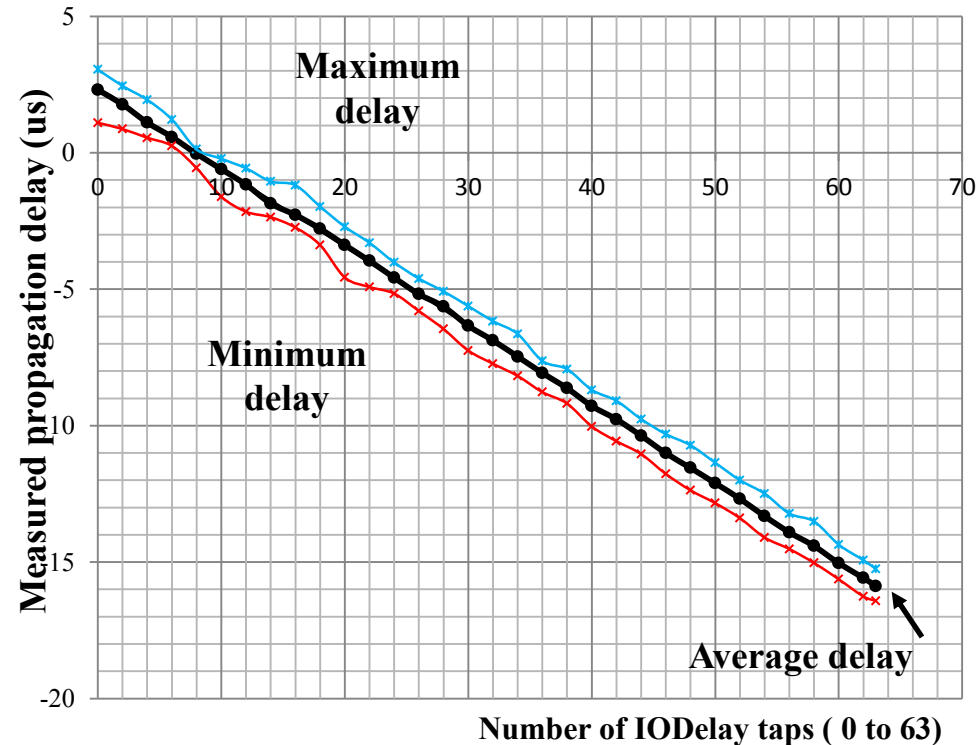
DDMTD

Using the same ML507 testbench, Clock1 is generated as a delayed (out of phase) Clock0 with a programmable delay multiple (0 to 63) of 78ps in addition to propagation delay.

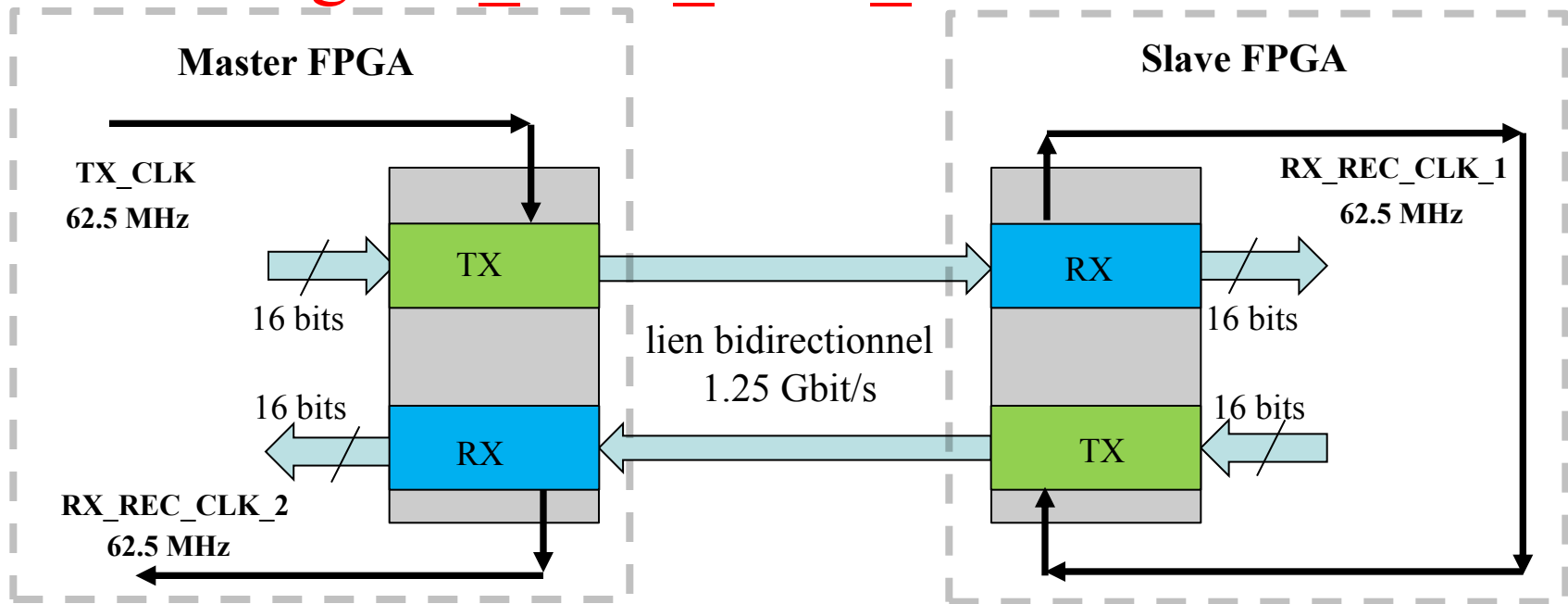


Delays amplified by a factor $1 / 0,0002688$ are more conveniently measured on output.

Estimated slope is $0,2888 \mu\text{s}$ per delay tap which is coherent after conversion with 78 ps per tap :
 $0,2888 * 0,0002688 * 1000000 = 77,65 \text{ ps}$.



Transmitting back from Slave to Master using RX_REC_CLK_1



➔ Solution simple : PMASETPHASE = '1'
ENPMAPHASEALIGN = '1'

Risque : augmentation du jitter.

➔ Autres solutions avec PLL externe

Measuring propagation delay over a 1.25 Gbps bidirectional data link

P.P.M. Jansweijer,^a H.Z. Peek,^a

^aNikhef, Science Park 105, 1098 XG Amsterdam, Netherlands



White Rabbit

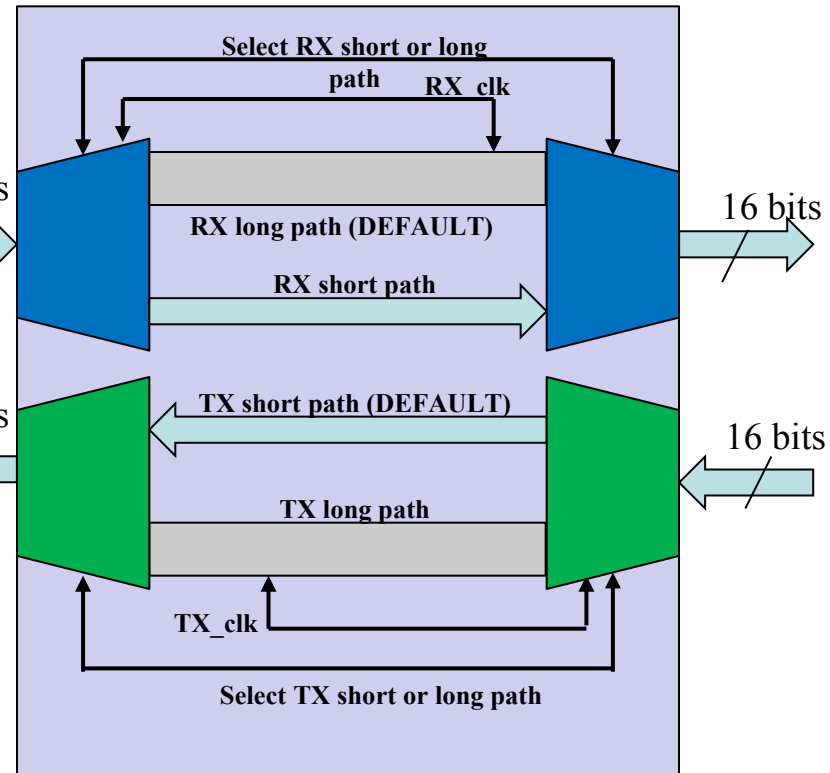
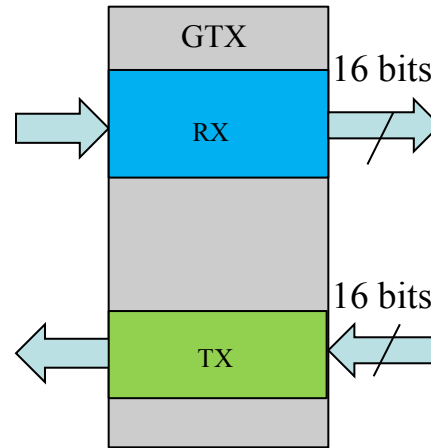
TX_CLK, RX_REC_CLK_1, RX_REC_CLK_2 sont trois horloges déphasées issues de la même source.

DDMTD pour mesurer le déphasage entre TX_CLK et RX_REC_CLK_2

Synchronous Commands

Incoming commands :

- commands detected inside RX default long path pipeline. Could use a specific K-char to tag commands and second byte to define 256 commands.
- switch to short path , remove command from RX pipeline and replace it with 16 bit idle.
- switch back to default long path within inter frame gap, which is then one idle longer than initial gap.



N.B. The data bus is pipelined along with several other signals (running disparity, char_is_K, etc.)

Outgoing commands :

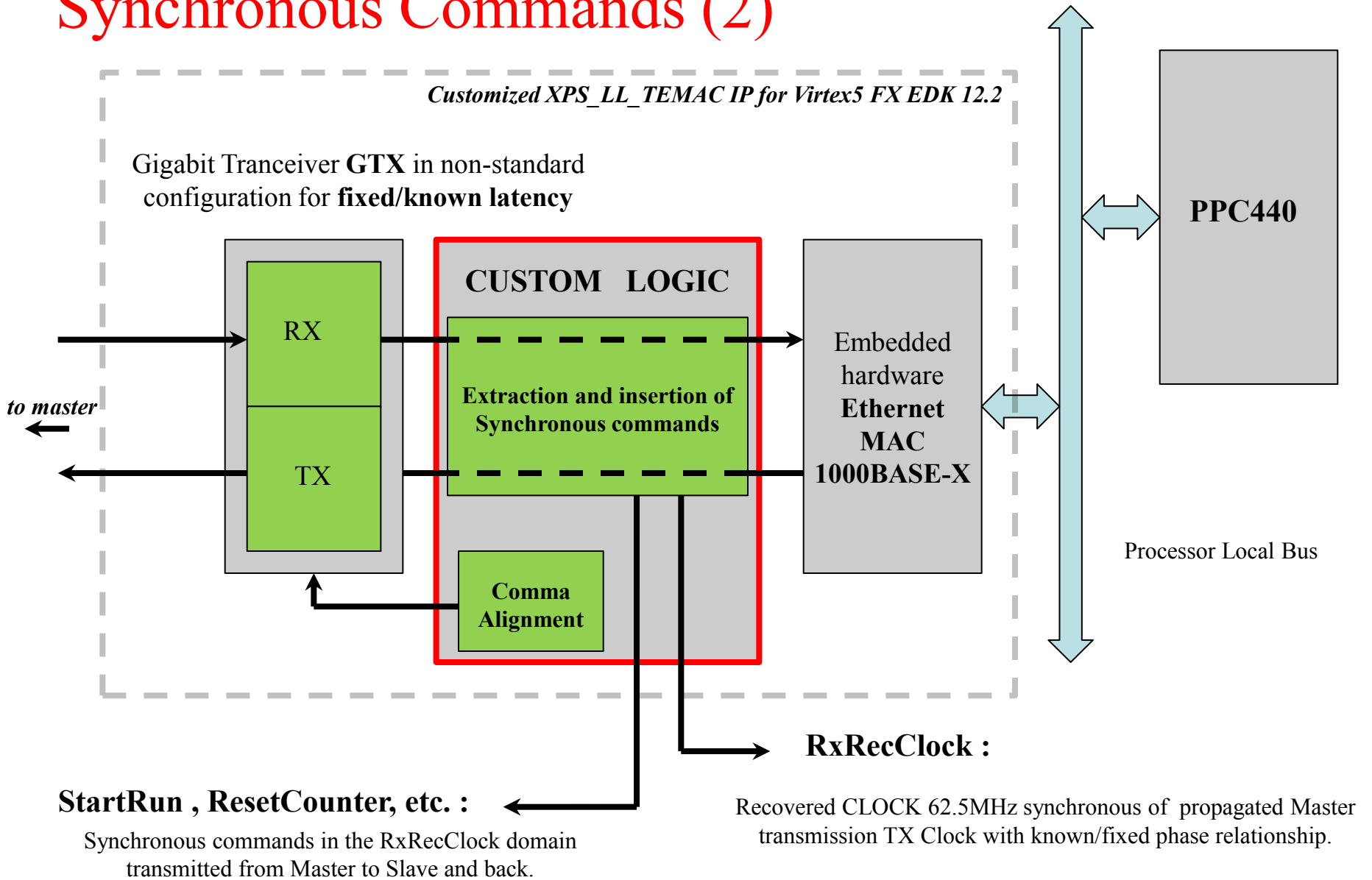
- command to be inserted is stored in long path pipeline
- switch to long path : move command out and pipeline Ethernet flow
- switch back to default short path within inter frame gap by removing one 16 bit idle.



Able to handle the necessary low rate of synchronous commands.

Commands inserted anywhere inside Ethernet flow without corrupting ethernet frames.

Synchronous Commands (2)



- 1000BASE-X Ethernet drivers for VxWorks6.3

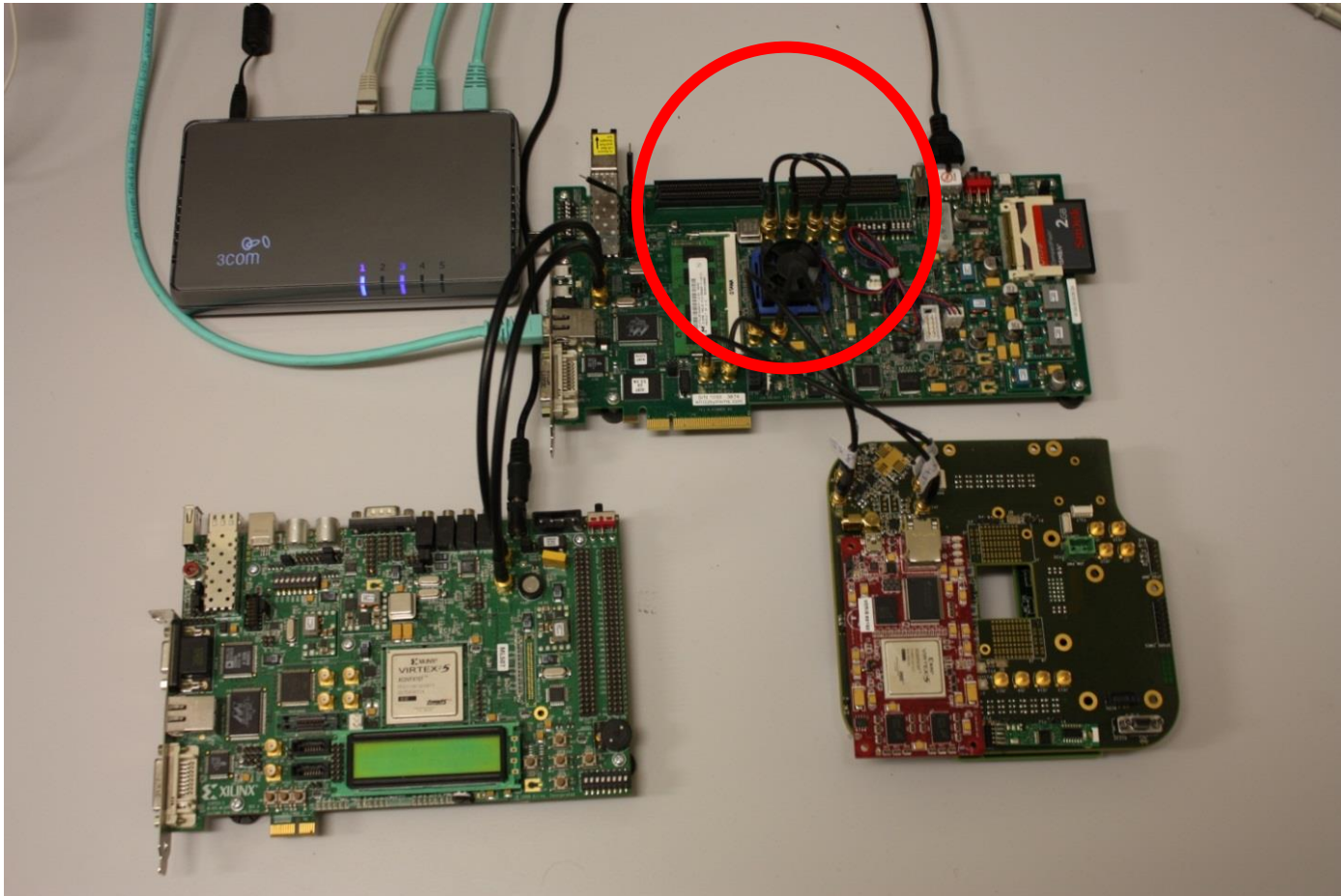
- logic insertion between TEMAC and GTX in the XPS_LL_TEMAC IP requires EDK12.2 (sources encrypted in prior versions!)

TDC pour mesure de latence AR

Commande synchrone émise vers le FPGA Escalve => “START” synchrone de TX_CLK

Commande synchrone recue (echo) par le FPGA Maitre => “STOP” synchrone de RX_REC_CLK_2

Ces impulsions sont combinées et sorties FPGA Maitre sur un seul lien LVDS vers l’entrée différentielle d’un TDC implémenté sur ce même FPGA.

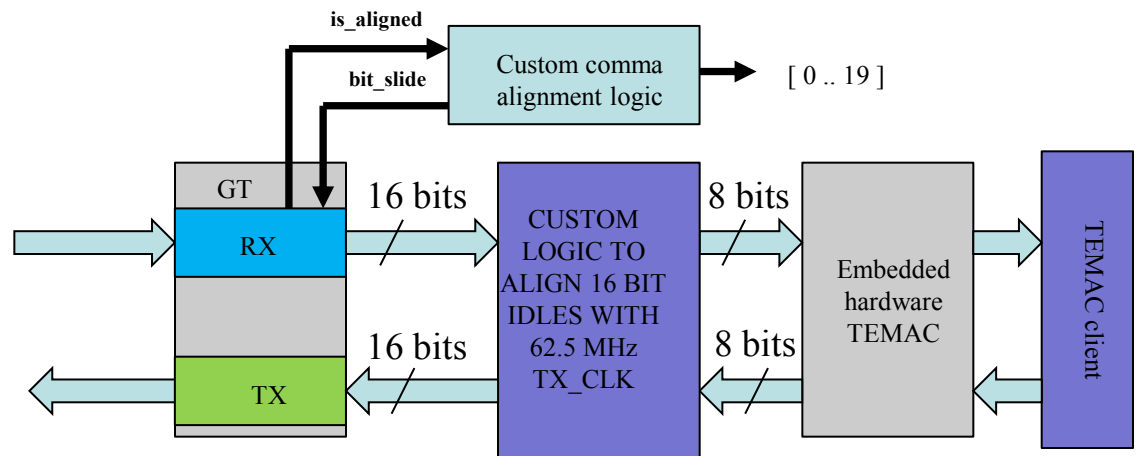


BITSLIDES et alignement manuel

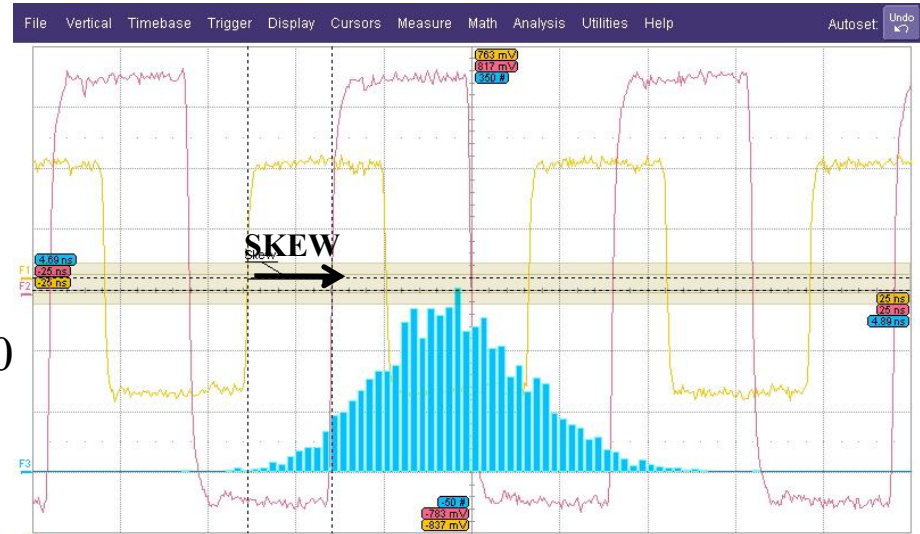
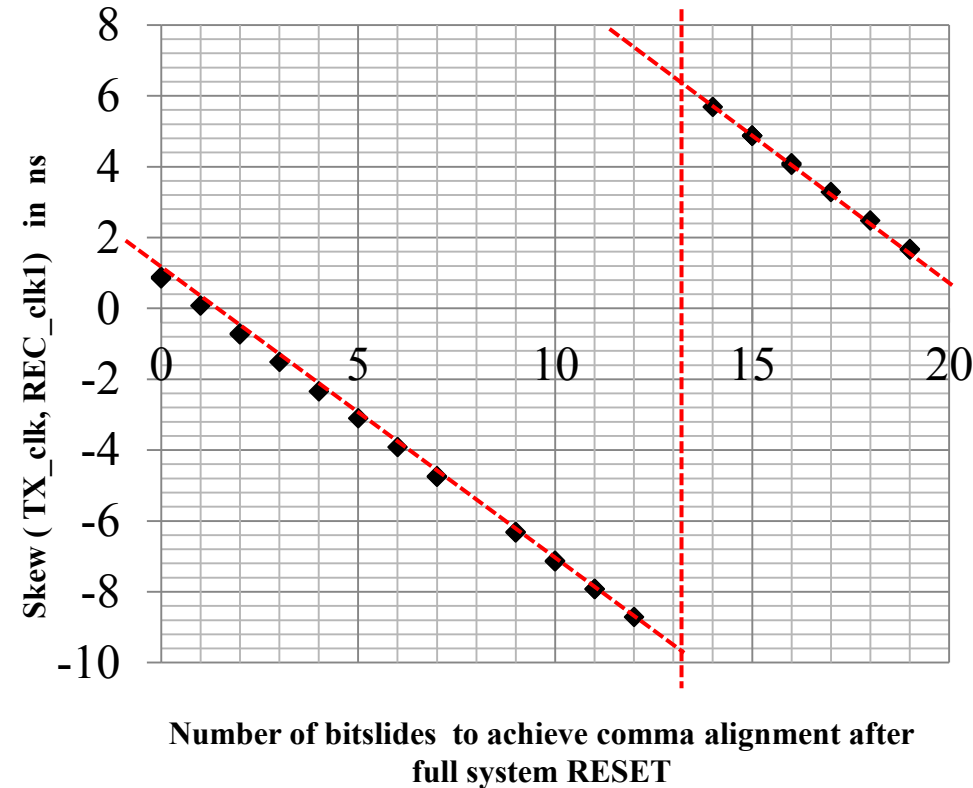
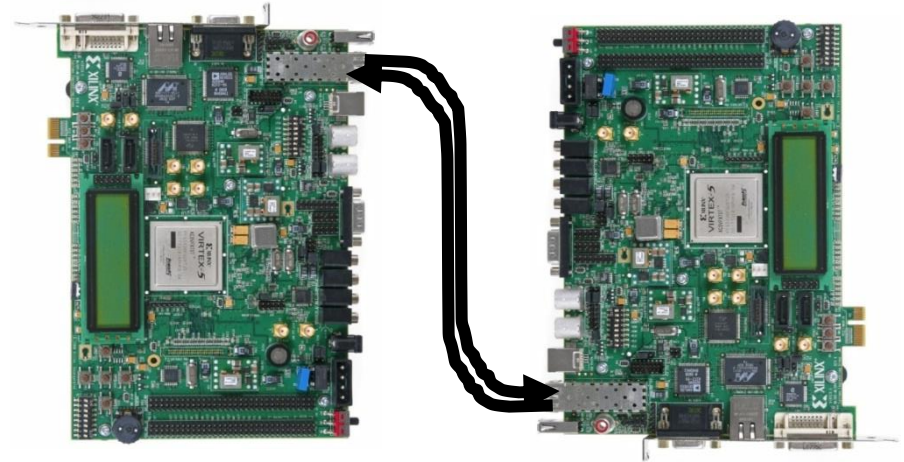
- Bus interne en 16 ou 20 bits de large. Interface utilisateur en 8, 16 ou 32 bits de large. Besoin de mots IDLES / COMMAS de 16 bits (20 bits après encodage) pour lever les indéterminations.
- Du fait de la serialisation / deserialisation de 20 bits, l'horloge reconstruite peut presenter 20 decalages possibles relativement à l'horloge propagée synchrone des données.
- Le GTX peut compenser ces décalages et resynchroniser les données parallèles et la RecClock.
- Désactiver l'alignement automatique. Trouver la bonne configuration de la logique d'alignement interne. Compter le nombre décalages de 0.8 ns nécessaires.
- GTX connecté au TEMAC qui génère par chance des IDLES 16 bits (K28.5+D5.6 ou K28.5+D16.2)
- Interface TEMAC coté GTX en 8 bits !
- Logique insérée entre TEMAC et GTX pour aligner correctement les IDLES sur l'horloge 62.5 MHz et pouvoir configurer le GTX en 16/20 bits. Necessary pour avoir une configuration valide de la logique interne d'alignement qui permette d'avoir une relation utilisable entre latence et nombre de Bitslides.

--Comma Detection and Alignment Attributes -----
ALIGN_COMMA_WORD_0 => 2,
COMMA_10B_ENABLE_0 => "1111111111",
COMMA_DOUBLE_0 => FALSE,
DEC_MCOMMA_DETECT_0 => TRUE,
DEC_PCOMMA_DETECT_0 => TRUE,
DEC_VALID_COMMA_ONLY_0 => TRUE,
MCOMMA_10B_VALUE_0 => "1010000011",
MCOMMA_DETECT_0 => TRUE,
PCOMMA_10B_VALUE_0 => "0101111100",
PCOMMA_DETECT_0 => TRUE,
RX_SLIDE_MODE_0 => "PCS", -- PMA

RXDATAWIDTH1 => "00",
TXDATAWIDTH0 => "00",
INTDATAWIDTH => '1',



BITSLIDES et alignement manuel : résultats

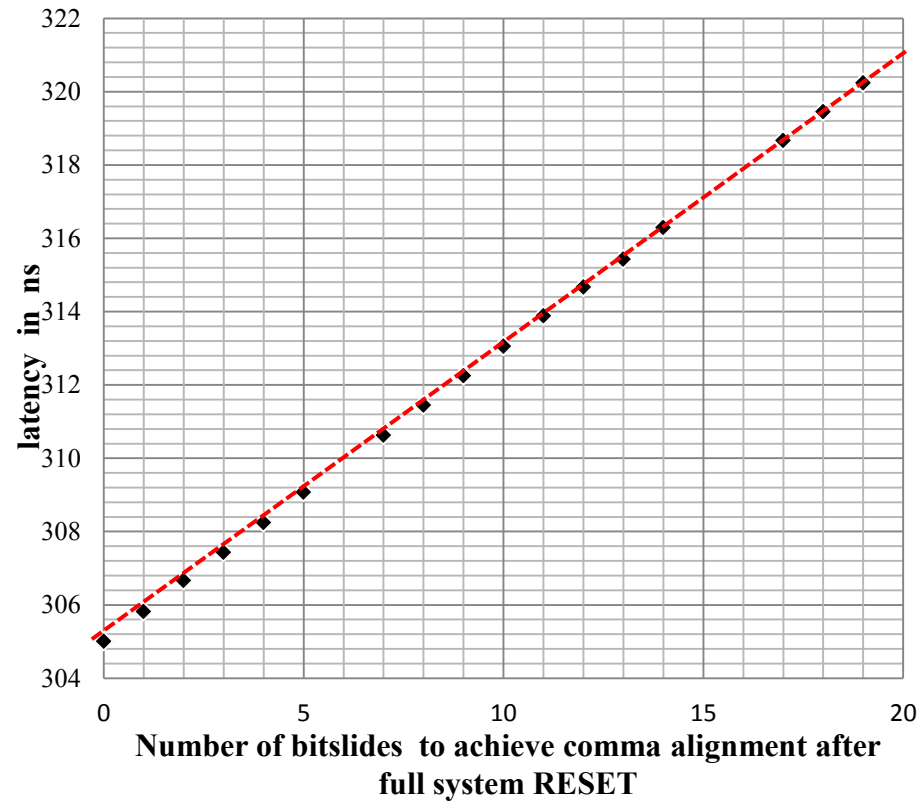
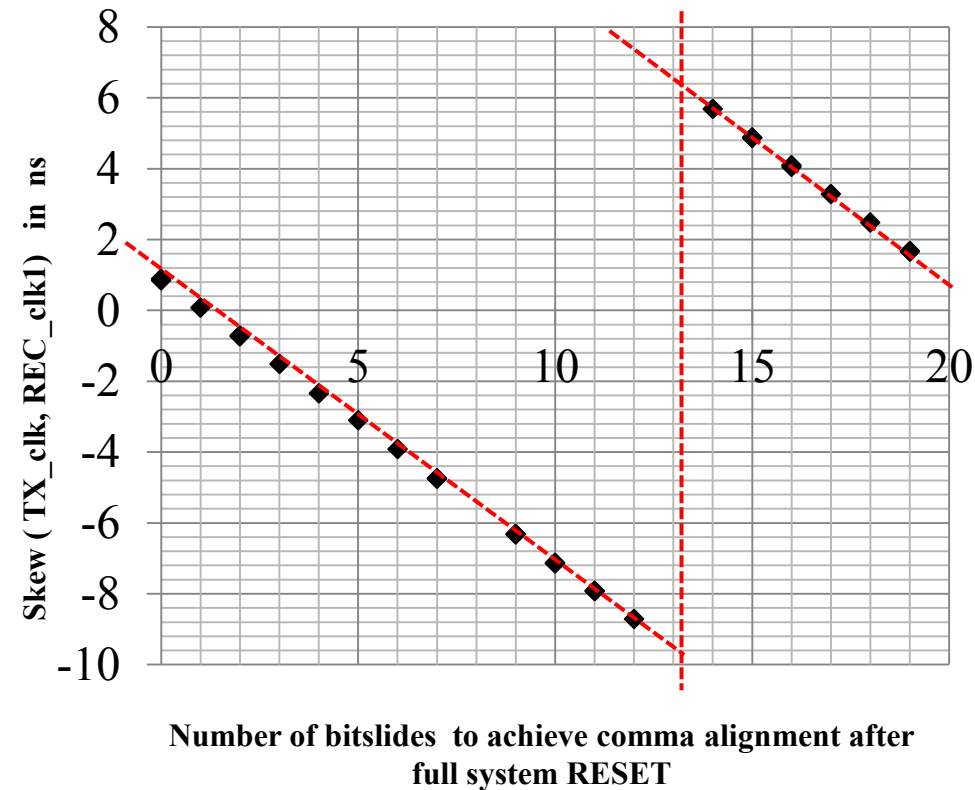
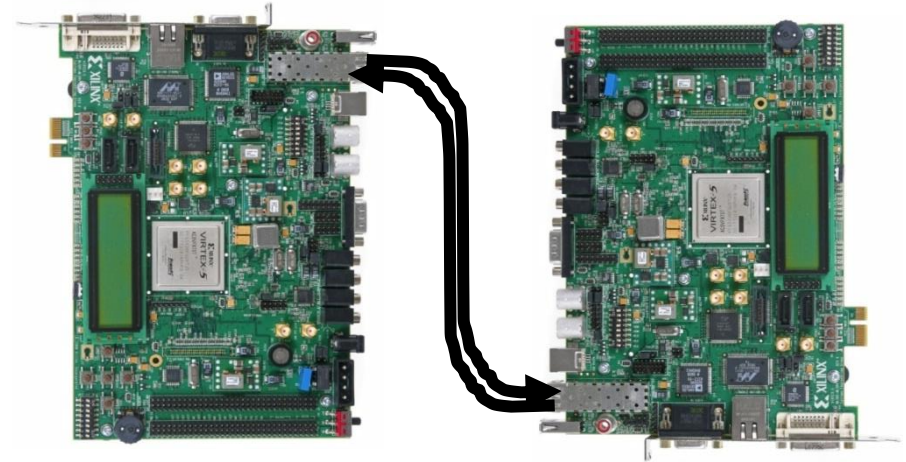


Histogram range < 200 ps

Histogram standard deviation < 25 ps

Sampling at 10 GS/s

BITSLIDES et alignement manuel : résultats



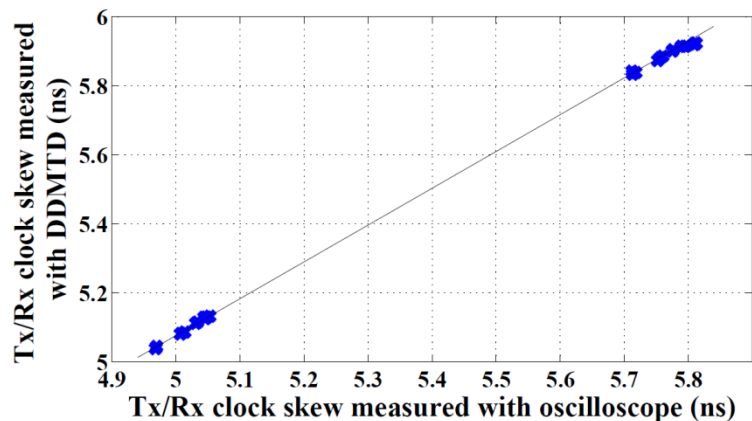
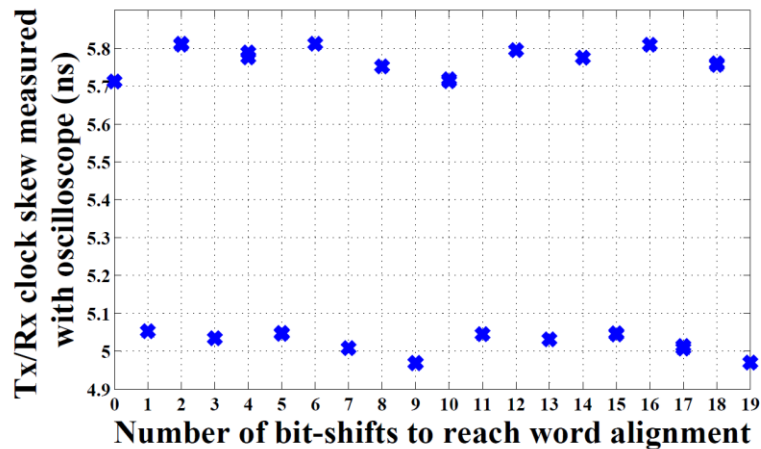
PCS vs PMA bitslides

The GTX can be configured to achieve manual bit sliding in the PCS or in the PMA.

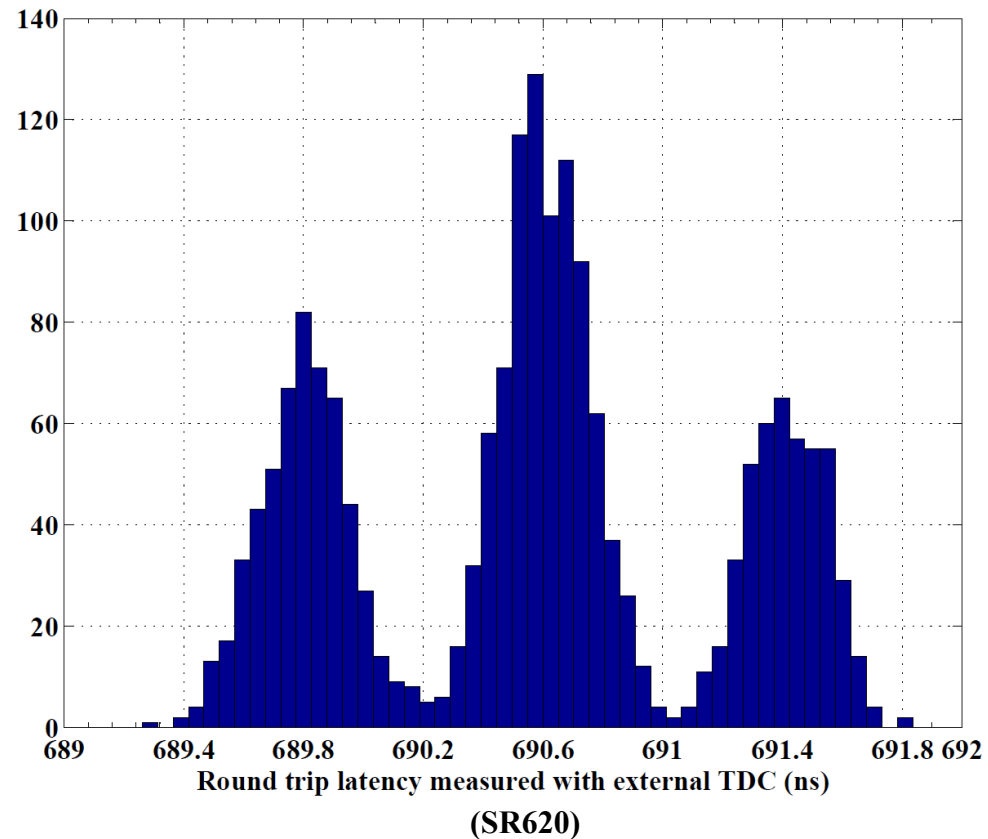
PCS : 20 different values of clock skew and latency associated to 20 different values of the number of bitslides (one way trip).

PMA : 2 different values of clock skew and latency which map to even or odd numbers of bitslides. (R. Giordano *et al.*)

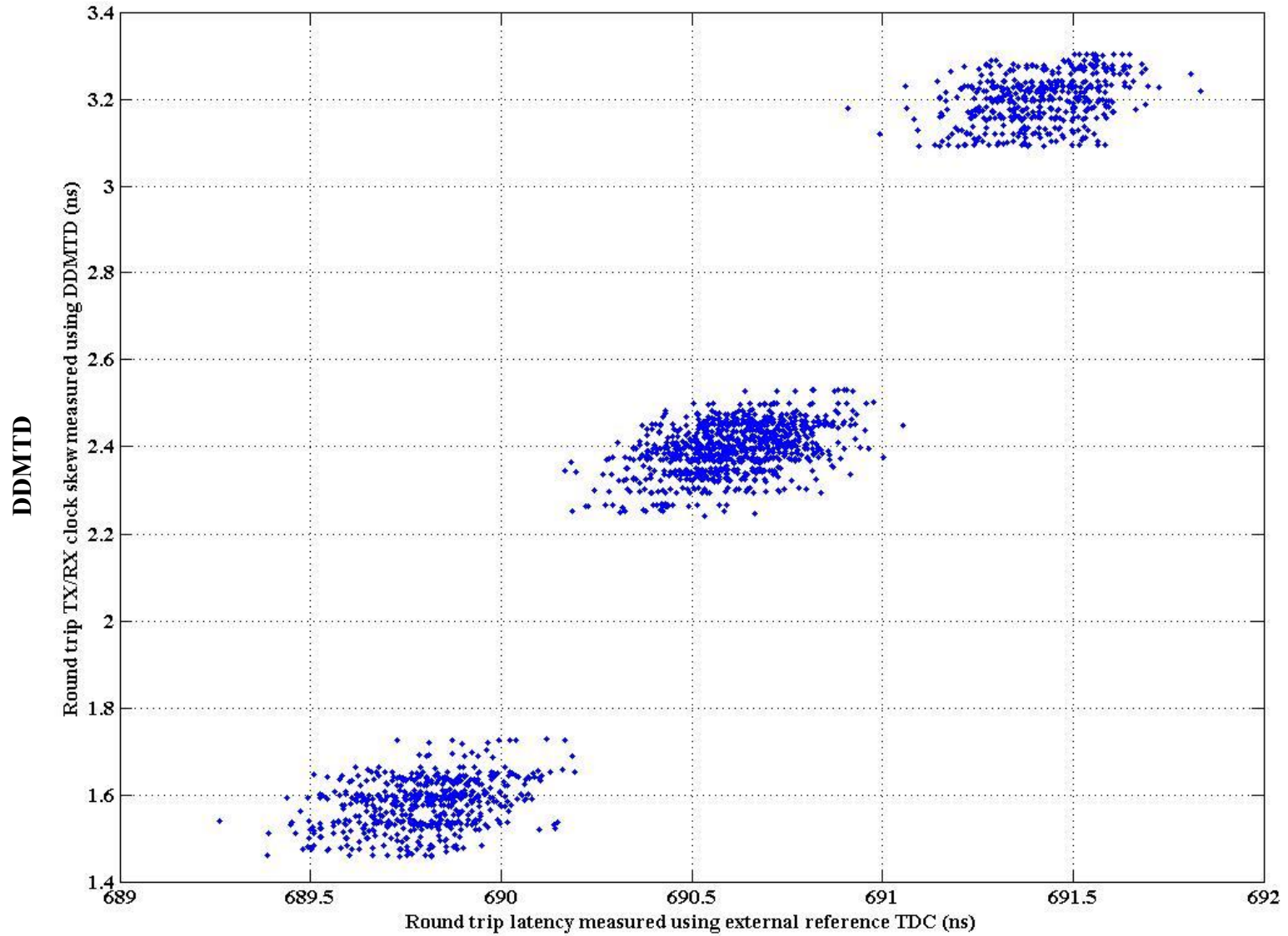
Aller simple



Aller et Retour

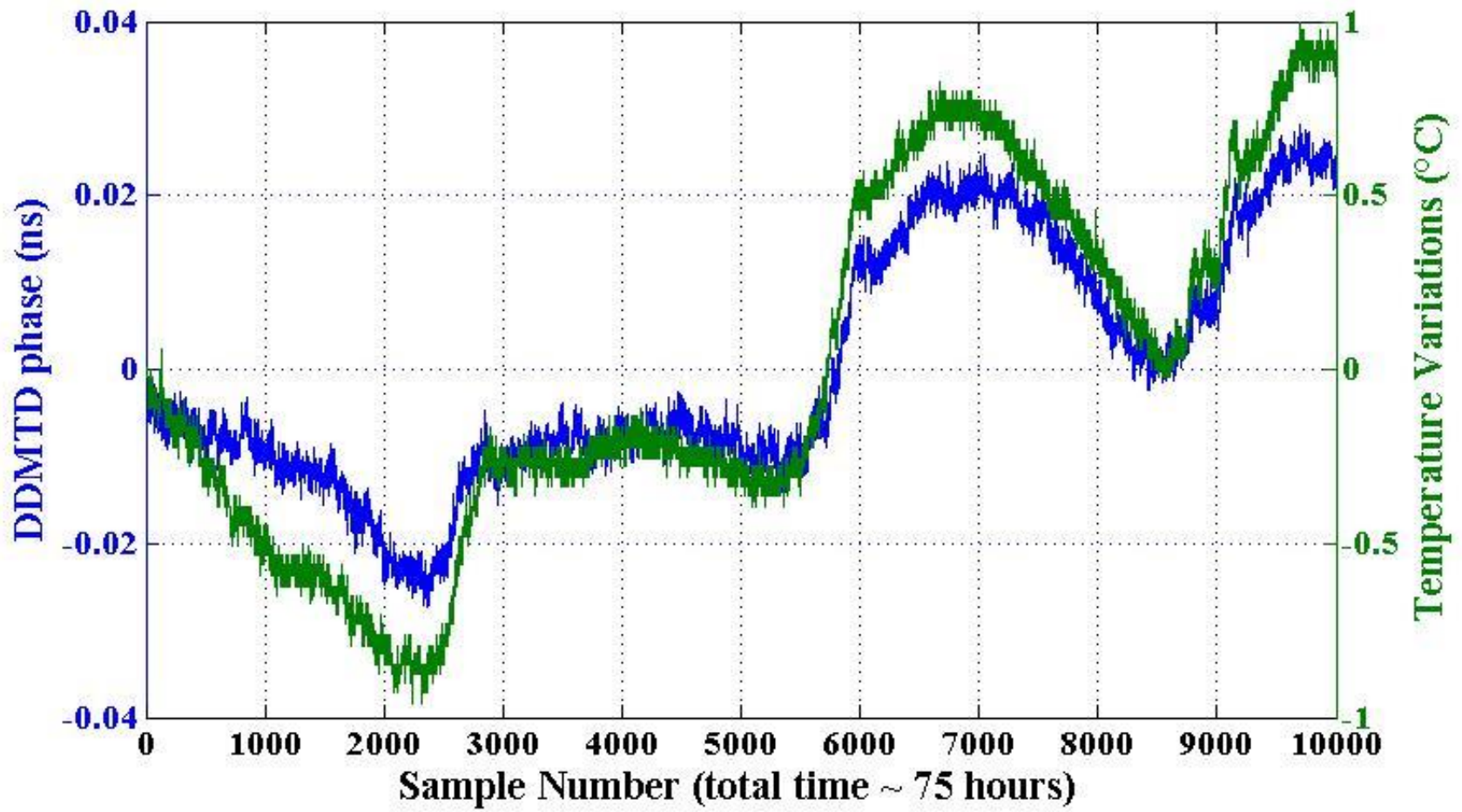


Alignement et Latence AR

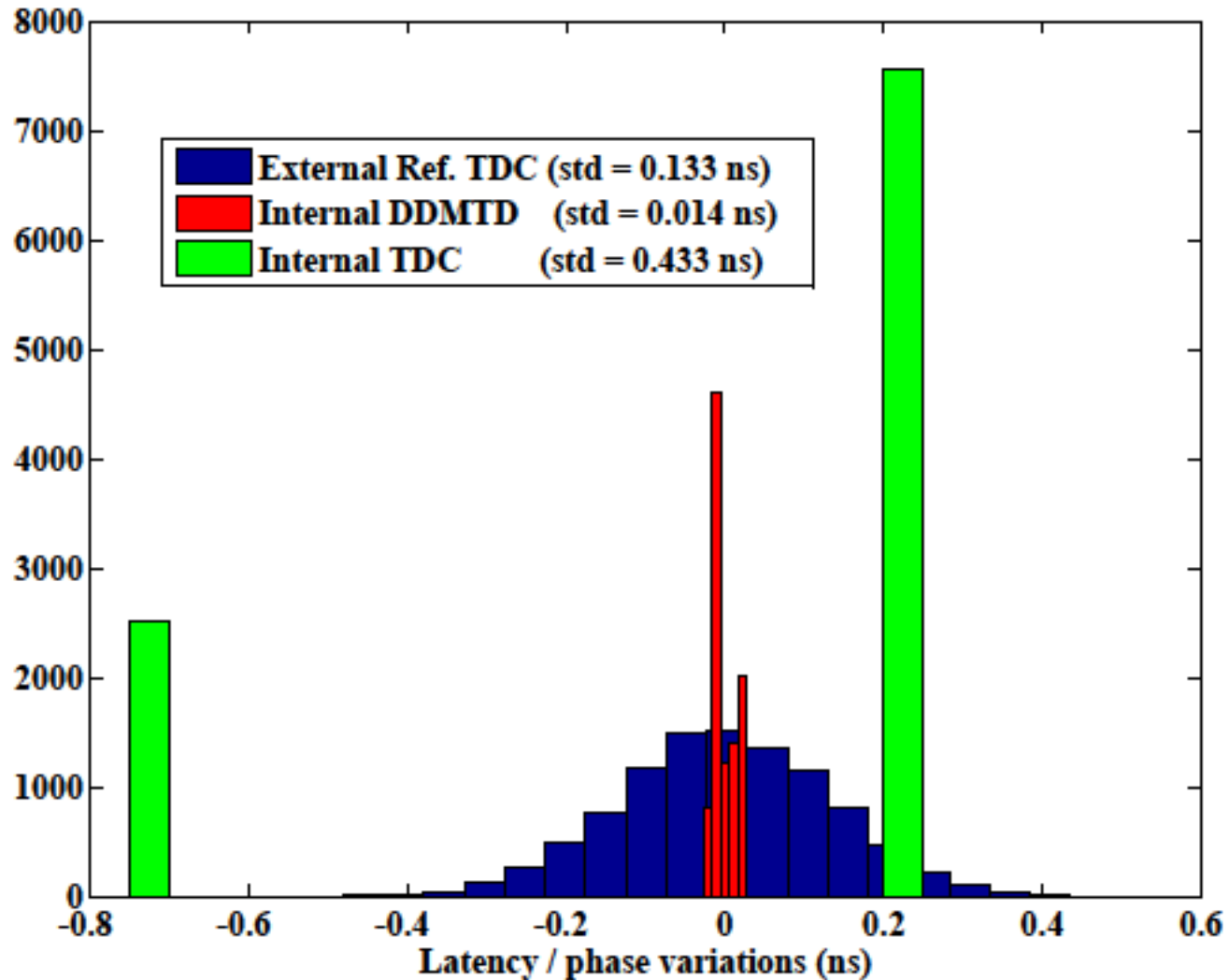


(SR620)

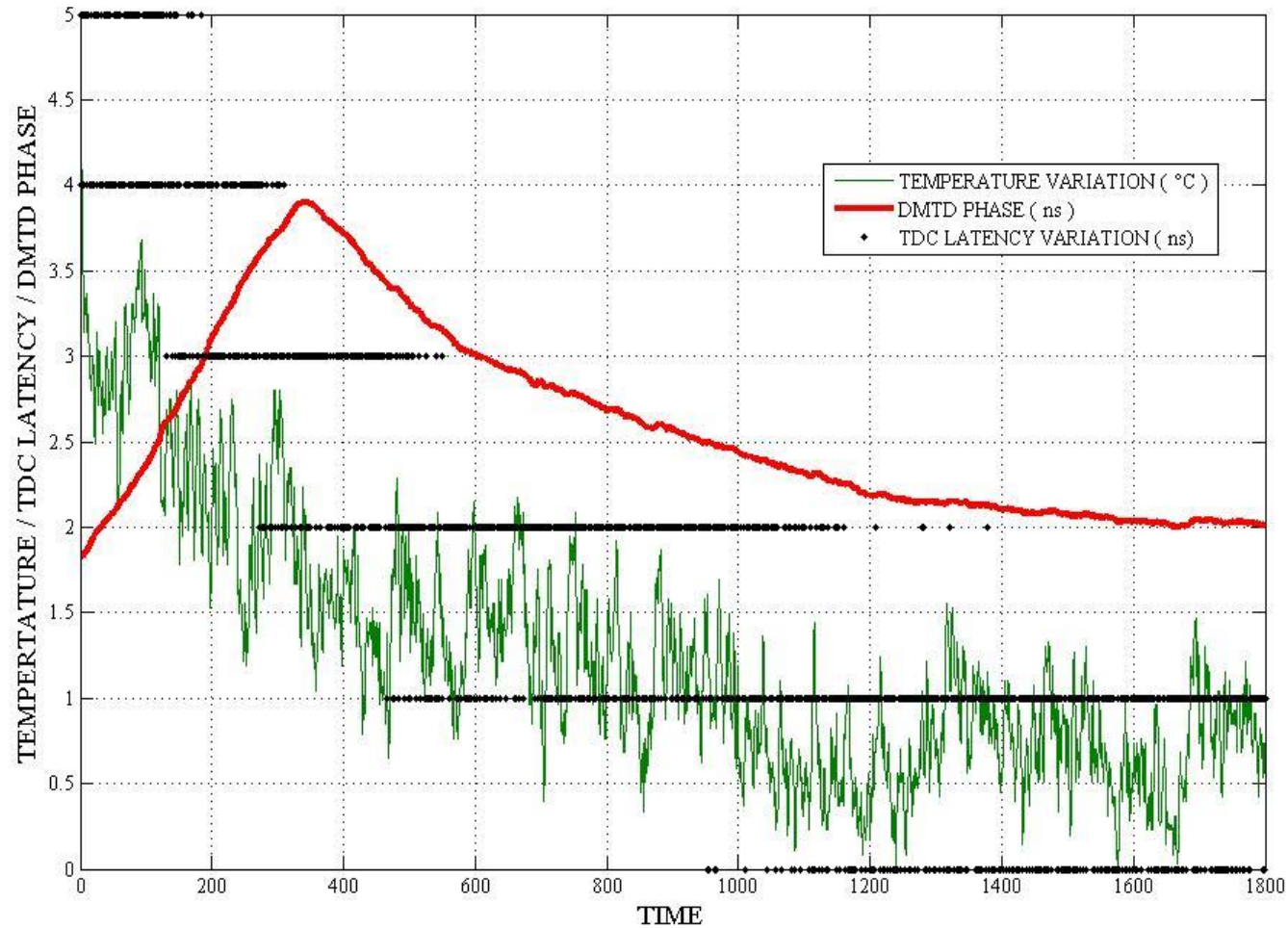
Variations de la Latence AR avec la Température



Variations de la Latence A/R avec la Température



Mesures de phase, latence et température



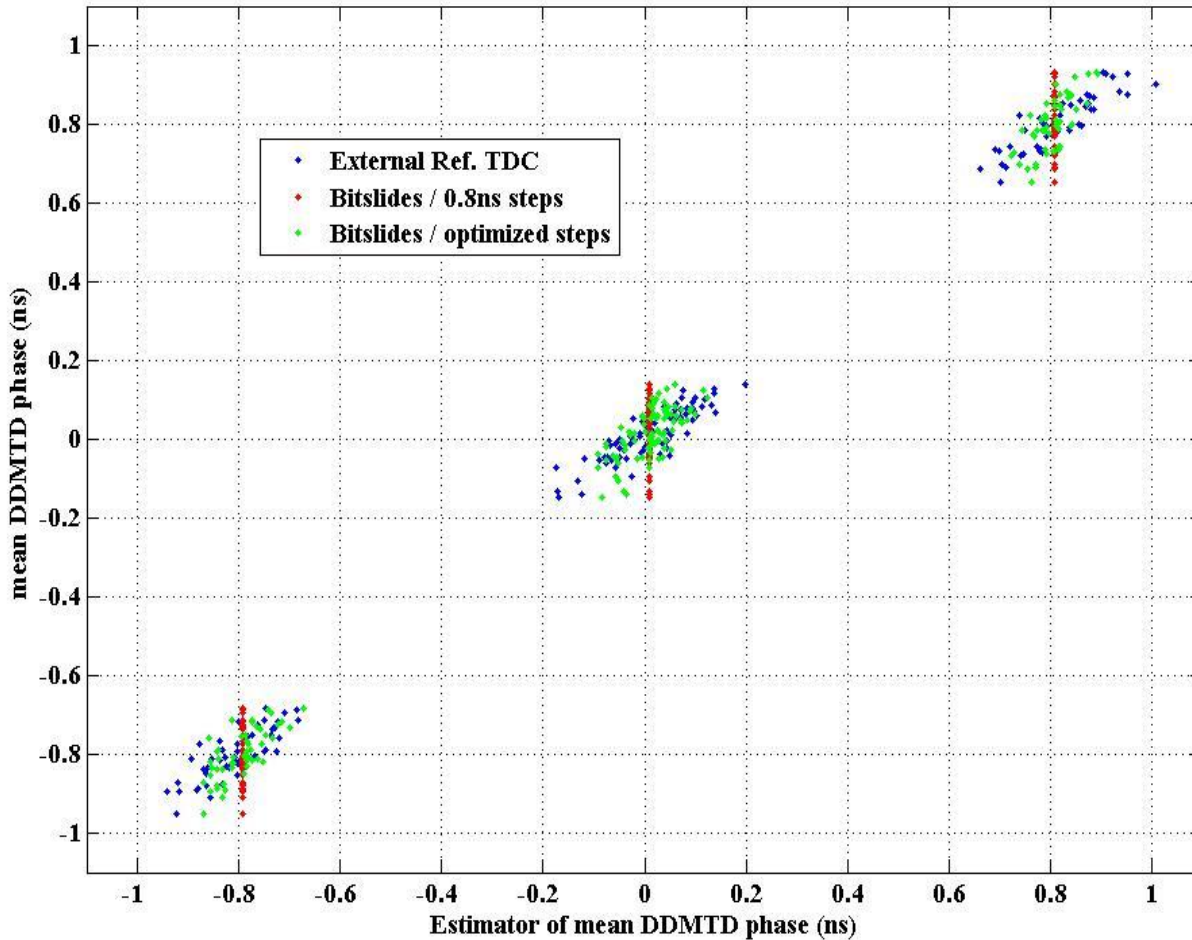
~ 10 h



Utiliser plusieurs horloge DDMTD déphasées pour mieux gérer la périodicité de la mesure de phase.

Mesures A/R et Calibration

- Probleme : attribuer à l'aller et au retour sa part de latence variable due à l'alignement.



```
MM = zeros(39,1);  
DD = zeros(39,1);  
AA = zeros(20,19);
```

```
for ii=1:20
```

```
    A_on = 1 * ( BITSLIDES_on == (ii - 1) );
```

```
    MM(ii) = A_on * DMTD';
```

```
    DD(ii) = A_on * A_on';
```

```
    for jj = 1:19
```

```
        A_off = 1 * (BITSLIDES_off == (jj) );
```

```
        AA(ii,jj) = A_on * A_off';
```

```
        MM(jj + 20) = A_off * DMTD';
```

```
        DD(jj + 20) = A_off * A_off';
```

```
    end
```

```
end
```

```
BB = diag(DD);
```

```
CC = [[zeros(20,20), AA] ; [AA', zeros(19,19)]];
```

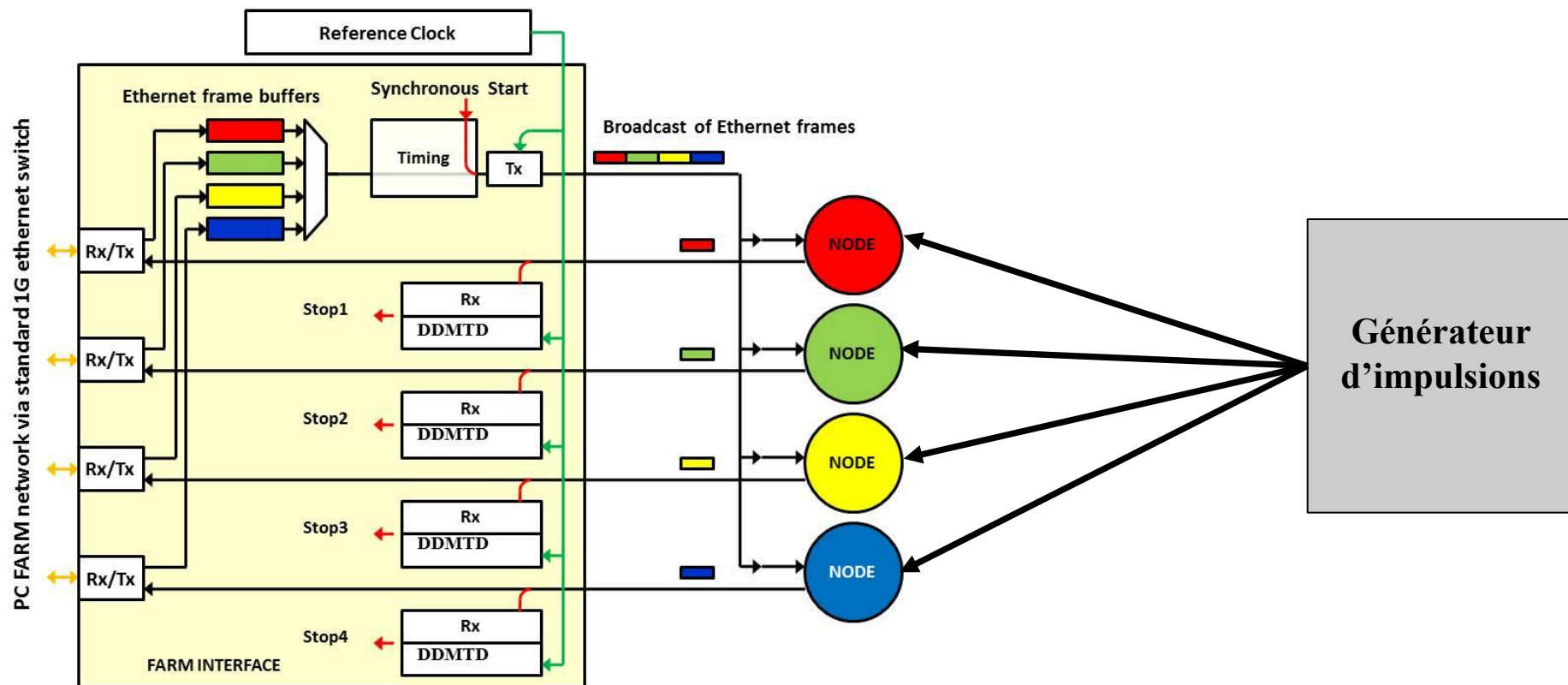
```
vv = inv(BB+CC) * MM;
```

Inversion d'un systeme lineaire au sens des moindres carrés

Mesures sur un aller et retour

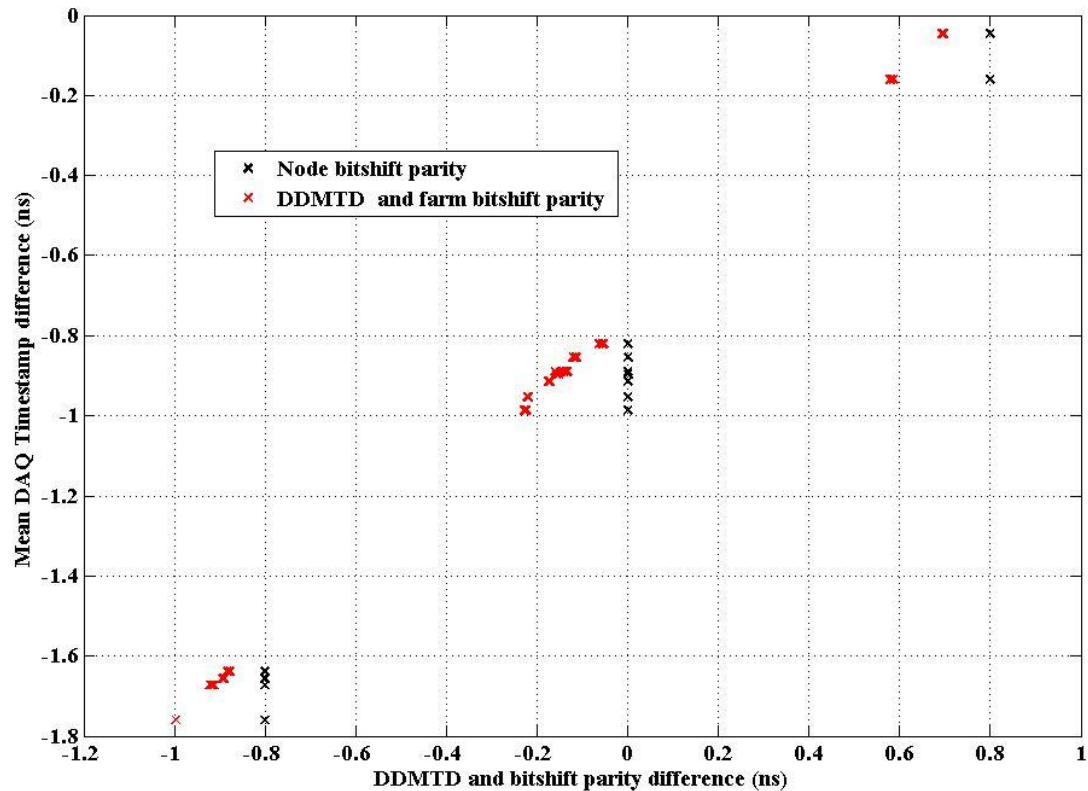
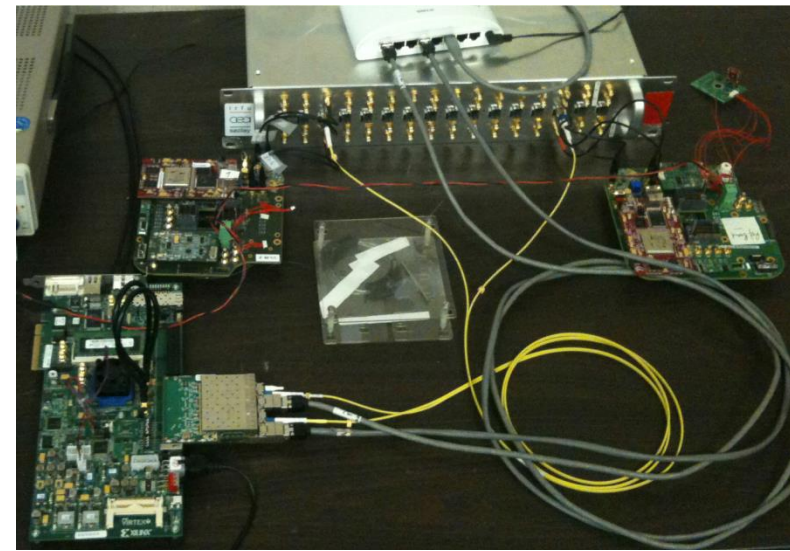
Conclusion

➔ L'objectif est de pouvoir remettre en temps les données issues de plusieurs nodes.



Conclusion

- Two slave nodes.
- GTX reset periodically on both Slaves and Master.
- Forces realignment on word boundaries and modifies Tx and RX latencies between Master and Nodes.
- Injected pulses are timestamped on each slave node.



Références

A readout system-on-chip for a cubic kilometer submarine neutrino telescope

H. Le Provost *et al.*, TWEPP 2011

Design and implementation of a nanosecond time-stamping readout system-on-chip for photo-detectors

S. Anvar *et al.*, Nuclear Instruments and Methods in Physics Research A, Volume 735, p. 587-595.

Time calibration of the ANTARES neutrino telescope

J.A. Aguilar *et al.*, Astroparticle Physics, 34, 2011

Measuring propagation delay over a 1.25 Gbps bidirectionnal data link

P. Jansweijer and H. Peek, Atlas note, 2010

PET system synchronization and timing resolution using high speed data links

R. Aliaga *et al.*, RT2010, Lisbonne.

Fixed latency, multigigabit serial links with Xilinx FPGAs

R. Giordano *et al.*, IEEE Nuclear Science, 2011

Precise time and frequency transfer in a White Rabbit network

Tomasz Wlostowski, Master Thesis

White Rabbit homepage : <http://ohwr.org/projects/white-rabbit/documents>



Highly accurate timestamping for ethernet based clock synchronization

P. Loschmidt *et al.*, Journal of Computer networks and Communication, 2012

FPGA to IBM Power Processor Interface Setup

I. Ouda *et al.*, Application Note, IBM Research Report

LVDS 4x Asynchronous Oversampling Using 7 Series FPGAs

Marc Defossez, Xilinx Application Note 523 (v1.0) April 6, 2012.