

ES102/PC5 : énoncé et corrigé

1) Inverseur CMOS en « grands et petits signaux »

Ci-contre sont présentées les caractéristiques $I_{DS}(V_{DS})$ d'un transistor nMOS, pour différentes valeurs de V_{GS} . Il s'agit d'un transistor de taille minimale dans une technologie CMOS 350nm (bien plus grossière que les technologies numériques de pointe), toujours exploitée, mais pour des circuits analogiques.

- 1a) Où se trouve la zone non passante sur la figure ?

Sur l'axe horizontal...

- 1b) $3.9 \times 8.85 \cdot 10^{-12} \times 0.35 \cdot 10^{-6} \times 0.5 \cdot 10^{-6} / 8 \cdot 10^{-9}$: ainsi est estimée à 0,75 fF la capacité de grille du transistor ci-contre. Expliquer ce calcul.

Il s'agit de la capacité entre grille et substrat du transistor. Bien que capacité MOS, elle est ici traitée comme une capacité classique et évaluée à $C = \epsilon_{SiO_2} \cdot \epsilon_0 \cdot W \cdot L / T_{ox}$ où $\epsilon_0 = 8,85 \cdot 10^{-12} \text{ F} \cdot \text{m}^{-1}$ est la permittivité diélectrique du vide et $\epsilon_{SiO_2} = 3,9$ la permittivité relative de la silice. L'épaisseur de silice est $T_{ox} = 8 \cdot 10^{-9} \text{ m}$, pour *oxide thickness*. La réalité est un peu plus complexe dans une capacité MOS, mais cette approximation suffit à notre démarche.

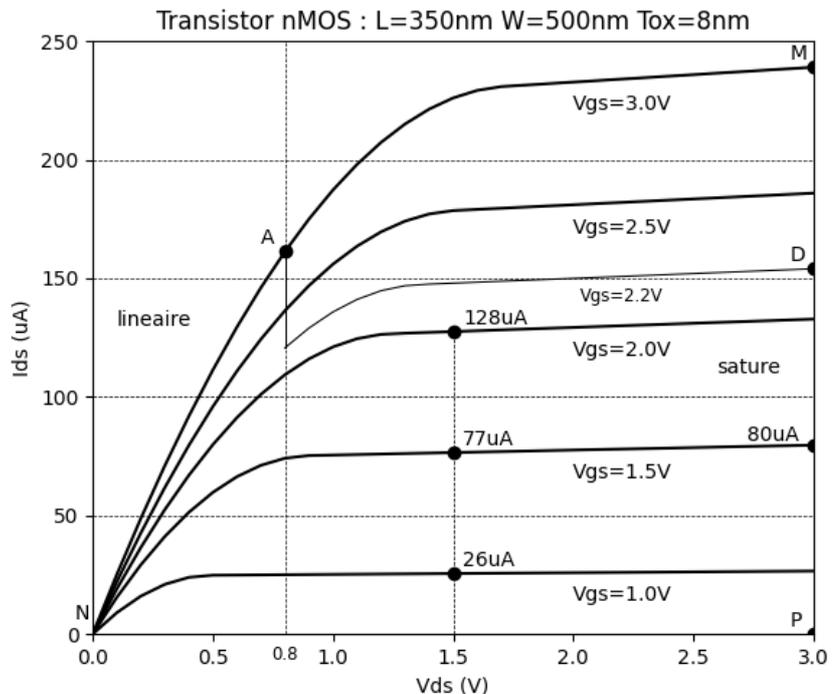
- 1c) La capacité d'entrée C_{inv} d'un inverseur minimal équilibré sera donc d'environ 3fF. Pourquoi ?

Transistors *n* et *p* de taille minimale ont exactement les mêmes dimensions, donc même capacité de grille. Mais les trous sont 3 fois moins mobiles que les électrons. Pour équilibrer l'inverseur, il faut donc tripler la largeur du pMOS, ce qui triple aussi sa capacité, à 2,25fF.

- 1d) On considère une chaîne de tels inverseurs, alimentés sous $V_{dd} = 3V$. Chacun demande donc le même délai pour transmettre (en l'inversant) une commutation, de son prédécesseur vers son successeur. Estimer ce délai τ_{inv} .

Produite par l'inverseur précédent, V_{in} va monter de 0V à 3V, rendant notre nMOS passant et notre pMOS non passant. A noter que celui-ci, juste avant la transition, ne fournit plus de courant, bien que passant, car ayant fini son travail de charge. Ainsi, globalement, il ne s'opposera que faiblement à l'action du nMOS et on le négligera.

On peut maintenant tenter d'imaginer la trajectoire suivie par le point (V_{DS}, I_{DS}) sur les caractéristiques ci-dessus, tandis que $V_{in} = V_{GS}$ s'élève. Cette trajectoire débute en bas à droite, en P, vers le haut. Dès que $V_{in} = V_{GS}$ est suffisamment haut, un courant I_{DS} significatif commence à décharger C_{out} et à faire baisser V_{DS} . Ainsi notre point va se mouvoir d'autant plus vite vers la



gauche qu'il est haut placé. La trajectoire ascendante s'incurve donc vers la gauche et finit par s'approcher de la caractéristique la plus élevée tandis que $V_{in}=V_{GS}$ tend vers V_{dd} .

Selon une première approximation radicale, on suppose qu'on monte instantanément en M, d'où $V_{GS}=3V$, ce qui nous place sur la caractéristique la plus haute. On estime alors le temps qu'il faut pour aller de M jusqu'à l'origine N. Mais ce temps est infini ! On se résigne alors à ne faire que 90% du chemin, par exemple, se contentant d'un bon '0' à défaut d'un '0' parfait. Même si, empiriquement, cela peut donner des résultats utilisables, ce n'est pas du tout satisfaisant sur un plan conceptuel.

Dans une chaîne d'inverseurs, une commutation en un point déclenche une *onde* de commutation sur les inverseurs situés en aval. Le délai utile est donc le temps de propagation de cette *onde* d'un inverseur à son successeur. Et pour le mesurer, il « suffit » de saisir l'écart de temps entre 2 situations équivalentes sur deux inverseurs successifs.

Par exemple, considérons notre nMOS parvenu à décharger la capacité de sortie jusqu'à 0,8V, comme suggéré par une ligne verticale présente sur la figure. Cela correspond à une décharge de 2,2V par rapport aux 3V initiaux (une commutation faite à 73%, donc). Pour l'inverseur prédécesseur, l'équivalent est une charge de 2,2V, partant de 0V, ce qui se traduit par du 2,2V à l'entrée de notre inverseur. Le délai de notre inverseur peut donc être mesuré par le temps qui s'écoule entre l'instant (de départ) où $V_{GS}=2,2V$ et l'instant (d'arrivée) où $V_{DS}=0,8V$.

Considérons la trajectoire réelle suivie par le couple (V_{DS}, I_{DS}) . On ne la connaît pas. Néanmoins on sait qu'elle franchira forcément la caractéristique $V_{GS}=2,2V$ à un instant donné, puisque V_{GS} passe de 0V à 3V. Appelons justement *ligne de départ* cette caractéristique (qui s'achève au point D sur la figure). Plus tard (par causalité), cette trajectoire franchira forcément aussi la ligne verticale $V_{DS}=0,8V$, puisque V_{DS} passe en conséquence de 3V à 0V. Appelons justement *ligne d'arrivée* ce segment de droite (qui s'achève au point A sur la figure).

Le délai de notre inverseur est donc le temps mis le long de cette trajectoire pour passer de la ligne de départ à la ligne d'arrivée. L'estimer précisément reste un problème ouvert, mais bien posé. Cependant, et c'est un progrès notable, on tient désormais un majorant évident. En effet, la décharge est d'au plus 2,2V, de D en A. D'autre part, la composante horizontale de la vitesse de parcours de la trajectoire est en chaque point proportionnelle à l'inverse de I_{DS} , représentée par l'ordonnée du point. Donc, plus on est haut, plus ça va vite. Or le courant est minoré entre les lignes de départ et d'arrivée : il apparaît qu'il va toujours rester au dessus de $150\mu A$, y compris côté gauche puisque la trajectoire « converge » vers la caractéristique la plus haute. D'où une estimation par excès de $3 \cdot 10^{-15} \times 2,2 / 150 \cdot 10^{-6} = 44ps$. Cela constitue notre réponse à la question posée : $\tau_{inv} < 44ps$.

Mais, en fait, la trajectoire réelle passera déjà à une certaine distance du point D, d'où un moindre travail de décharge à réaliser. Ou bien elle s'aventurera bien au dessus de $150\mu A$. On se hasarde donc à réduire de 25%, à la louche, notre estimation, soit $\tau_{inv} \approx 33ps$, une valeur sympathique pour la question suivante. Certains seront outrés par les libertés que nous prenons ainsi tout à coup, après tant de rigueur auparavant. Mais nous avons maintenant juste besoin d'une valeur plausible pour traiter la question suivante. Pour plus de précision, il faudrait recourir à un simulateur en temps continu (à la Spice), avec des transistors fidèlement modélisés.

Remarque : on aurait pu positionner A et D différemment pourvu qu'ils restent équivalents en termes de niveau de commutation. Mais le choix fait ci-dessus permet de bien maximiser le minorant sur le courant, ce qui est intéressant pour avoir la plus petite estimée par excès de τ_{inv} .

Autre remarque : les esprits curieux trouveront matière à aller plus loin dans la caractérisation de la trajectoire, mais c'est au-delà de nos objectifs pédagogiques.

- 1e) La question précédente donne une idée des délais combinatoires élémentaires pour la technologie considérée. Si l'on reboucle une chaîne de 3 tels inverseurs, que se passe-t-il ?

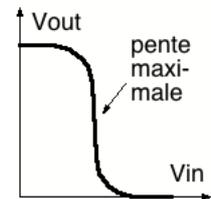
On obtient un oscillateur dit « en anneau » (situation d'instabilité évoquée au début du CM5). Si chaque inverseur met 33ps à faire commuter le suivant, il faut 99ps pour faire le tour complet,

ce qui représente une demi-période de l'oscillation globale, d'où une fréquence d'oscillation de l'ordre de 5GHz. Avec les technologies les plus récentes, on dépasserait 100GHz.

- 1f) En pratique, pour disposer d'un signal d'horloge, on se sert (presque) toujours d'un quartz externe et non d'un oscillateur en anneau interne. Devinez pour quelles raisons.

La fréquence d'oscillation d'un quartz varie très peu en fonction des conditions d'utilisation, contrairement aux caractéristiques de fonctionnement d'un circuit intégré. Premièrement, la température de fonctionnement T , qui varie facilement de 300 à 350K et au-delà, modifie significativement l'énergie élémentaire kT et donc le comportement physique du Silicium et des transistors. Deuxièmement, la tension d'alimentation peut elle-même varier. Troisièmement, des transistors pourtant fabriqués identiquement présentent nativement une certaine variabilité d'un individu à l'autre, d'autant plus lorsqu'ils sont de taille minimale.

On s'intéresse désormais aux aspects *petits signaux* de l'inverseur, en particulier à son gain, c'est-à-dire à la pente maximale de sa caractéristique $V_{out}(V_{in})$. On suppose le transistor nMOS ci-dessus associé à un transistor pMOS de « force » identique, en fait aux caractéristiques symétriques, c'est-à-dire tel que : $V_{GSp} = -V_{GSn}$ et $V_{DSp} = -V_{DSn} \Rightarrow I_{DSp} = -I_{DSn}$. Par symétrie, $V_{out} = V_{in} = 1,5V$ sera un point de fonctionnement de l'inverseur considéré et c'est en ce point que sera observée la pente maximale. Pour la calculer, il faut connaître la transconductance $g_m = \partial I_{DS} / \partial V_{GS}$ et la conductance de sortie $g_o = \partial I_{DS} / \partial V_{DS}$ du transistor nMOS au point considéré. En fait, c'est plutôt l'inverse de g_o , la résistance de sortie R_o , qui sera utilisée, pour simplifier les équations. Il faudra aussi utiliser ces grandeurs pour le transistor pMOS. On les distinguera par un indice n ou un indice p .



- 1g) Estimer les grandeurs g_{mn} et R_{on} au point de fonctionnement $(V_{GS}, V_{DS}) = (1,5V, 1,5V)$.

En $V_{DS} = 1,5V$, I_{DS} passe de 26 à 128 μA lorsque V_{GS} passe de 1.0 à 2.0V. La valeur de $g_{mn} = \partial I_{DS} / \partial V_{GS}$ en $V_{GS} = 1,5V$ peut ainsi être estimée par la pente moyenne $102 \mu A / 1V \approx 10^{-4} \Omega^{-1}$.

En ce même point, le nMOS est en mode saturé, ainsi que pour les valeurs plus grandes de V_{DS} . La valeur de g_{on} peut être estimée par la pente de la caractéristique sur la moitié droite, soit $(80-77) \mu A / (1,5V)$. D'où $R_{on} = 1/g_{on} \approx 500k\Omega$.

- 1h) Calculer analytiquement le gain de l'inverseur CMOS considéré autour d'un certain V_{in} , puis l'estimer en $V_{in} = 1,5V$, en se souvenant que, par symétrie, $g_{mp} = g_{mn}$ et $R_{op} = R_{on}$ en ce point.

Etant donné la structure de l'inverseur, on a : $V_{GSn} = V_{in}$; $V_{SGp} = V_{dd} - V_{in}$; $V_{DSn} = V_{out}$; $V_{SDp} = V_{dd} - V_{out}$. Ces grandeurs ont toutes été choisies positives, en ordonnant S, G et D de façon adéquate.

Considérons de petites variations dV_{in} et dV_{out} de V_{in} et V_{out} et exprimons leurs effets en courant. Côté n , $dV_{GSn} = dV_{in}$ et $dV_{DSn} = dV_{out}$, d'où :

$$dI_{DSn} = g_{mn} \cdot dV_{in} + R_{on}^{-1} \cdot dV_{out}$$

Côté p , $dV_{SGp} = -dV_{in}$ et $dV_{SDp} = -dV_{out}$, d'où :

$$dI_{SDp} = -g_{mp} \cdot dV_{in} - R_{op}^{-1} \cdot dV_{out}$$

Mais $I_{DSn} = I_{SDp}$ sinon il y aurait accumulation de charge, d'où :

$$g_{mn} \cdot dV_{in} + R_{on}^{-1} \cdot dV_{out} = -g_{mp} \cdot dV_{in} - R_{op}^{-1} \cdot dV_{out}$$

D'où finalement :

$$dV_{out} / dV_{in} = -(g_{mn} + g_{mp}) \cdot (R_{on} // R_{op})$$

où $(R_{on} // R_{op})$ est la résistance équivalente à R_{on} et R_{op} en parallèle. Le gain s'avère être un produit de deux facteurs. Le premier, $g_{mn} + g_{mp}$, est une conductance : c'est le coefficient de

proportionnalité entre une petite variation de tension d'entrée et le déséquilibre en courant qu'elle provoque. Le deuxième, $R_{on} // R_{op}$, est une résistance : c'est le coefficient de proportionnalité entre le déséquilibre en courant et la variation de tension de sortie qu'il provoque. Vues de la sortie, R_{on} et R_{op} apparaissent effectivement en parallèle.

Dans le cas symétrique considéré, on a $g_{mn}=g_{mp}=g_m$ et $R_{on}=R_{op}=R_o$, d'où un gain de $-g_m R_o$. Or $g_m \approx 10^{-4} \Omega^{-1}$ et $R_o \approx 500 \text{k}\Omega$, d'où un gain d'environ 50 (signe moins omis). Cela signifie que, à proximité de 1.5V, augmenter V_{in} de 1mV fait diminuer V_{out} de 50mV. La caractéristique de transfert de l'inverseur CMOS est donc très raide en son centre, lui valant toutes ses vertus. La raideur est moindre dans les technologies les plus récentes, mais demeure convenable.

Pour de véritables applications d'amplification, il est possible de faire bien mieux avec des transistors plus longs et larges — donc plus encombrants et capacitifs malheureusement — ce qui augmente R_o . Et pour des gains plus élevés, on peut chaîner plusieurs inverseurs : les gains se multiplieront (mais des problèmes de stabilité apparaîtront en boucle fermée...).

2) Premiers montages séquentiels

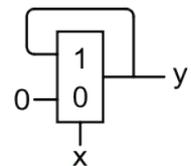
La logique séquentielle diffère de la logique combinatoire en ce que les sorties présentent des valeurs qui dépendent de l'historique des entrées, grâce à des montages en boucle fermée.

L'électronique numérique moderne repose sur la logique séquentielle dite *synchrone* (et son élément fondamental, la bascule D), séquencée par un signal d'horloge. Pour mieux l'introduire, on considère d'abord ici un montage séquentiel *asynchrone*, donc sans horloge.

- 2a) Comment se comporte la sortie y du montage ci-contre en fonction de l'historique de son entrée x ? Pour enrichir l'analyse, on peut remplacer la porte AND par son BDD (multiplexeur sur cofacteurs négatif et positif).



Mettre un circuit combinatoire en boucle fermée provoque presque toujours des effets indésirables. Mais celui-ci présente un état stable correspondant à $y=0$. En effet, quelle que soit la valeur de y à un instant donné, il suffit que x passe à 0 un bref instant pour faire passer y à 0 lui aussi (s'il n'y était déjà). Ce 0 en sortie s'impose sur la deuxième entrée et impose en retour à la sortie y de rester désormais à 0 quelles que soient les valeurs futures de x . En résumé, à supposer que y soit à 1 initialement (grâce des moyens d'initialisation restant à définir), la première chute à 0 de x fera chuter y définitivement (enfin, jusqu'à une future réinitialisation de y à 1). Le montage ci-dessus peut par exemple servir de détecteur de (micro)coupures sur un signal x devant normalement rester à 1 en permanence. Remplacer la porte AND par son BDD fait apparaître encore plus clairement la valeur de y maintenue à l'identique si $x=1$, mais mise à 0 si $x=0$.

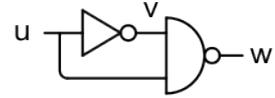


- 2b) La porte AND n'est qu'une abstraction de la réalité. Mais, ce niveau d'abstraction (le niveau porte, en l'occurrence) restitue-t-il correctement les propriétés de stabilité tenues pour acquises ci-dessus ? Méfiant, on peut en douter. S'en assurer en descendant au niveau *transistors*.

Descendre au niveau transistor nous oblige à nous remémorer qu'une porte AND est implantée par mise en série d'une NAND et d'un inverseur, dont les structures en transistors nous sont désormais familières. Il apparaît alors clairement que, lorsque $x=1$, on a l'équivalent de 2 inverseurs rebouclés l'un sur l'autre. Il s'agit du montage bistable étudié en cours : nous voilà rassurés ! Ceci garantit que non seulement $y=0$ est un état stable du montage, mais aussi $y=1$ lorsque $x=1$. Ainsi, $y=1$ demeurera tant que x ne passera pas 0. Par contre, il suffit d'un très bref passage à 0 de x pour faire basculer y , typiquement aussi bref que le temps de réponse d'un inverseur. On laisse de côté pour l'instant la question de la réinitialisation (qui nous emmènerait ici vers la bascule dite R'S') pour la reprendre plus loin en contexte synchrone.

Le montage ci-dessus est ultra-réactif : il suffit d'une très brève impulsion sur x pour le faire basculer. Cette réactivité est le propre des circuits asynchrones. Mais la réactivité peut être contre-productive sur les transitions parasites d'un signal d'entrée...

- 2c) D'un point de vue logique, le montage ci-contre, d'entrée u , fournit toujours la sortie $w=1$. D'où un caractère peu réaliste, sur lequel on ne se méprendra pas cependant : il s'agit d'un cas d'école, c'est-à-dire le plus simple possible, pour illustrer un problème d'ordre général. Prenant en compte le temps de réponse de l'inverseur, la sortie w demeure-t-elle effectivement à 1 lorsque l'entrée u varie ?

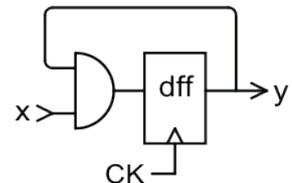


D'un point de vue logique, on a toujours $v=u$, d'où $w=1$. Pour avoir $w=0$, il faudrait que $u=v=1$. Or ceci se produit transitoirement lorsque u passe de 0 à 1, car v n'effectue la transition opposée qu'environ τ_{inv} plus tard. Sur la sortie w , on observe donc une brève impulsion vers le bas, qui aurait typiquement juste la durée nécessaire pour parvenir à faire basculer notre « détecteur de microcoupures » précédent. Cette irrégularité est qualifiée d'aléa (*hazard*). Elle se produit lorsqu'une même entrée affecte différemment la sortie selon différents chemins. Une transition sur cette entrée se propage le long des différents chemins pour atteindre la sortie à des instants différents et y provoquer des retournements de situation en cascade. On parle de *race condition* en anglais (pour situation de compétition). De tels problèmes se retrouvent aussi en informatique parallèle ou distribuée.

L'exemple précédent souligne que la logique combinatoire est un monde agité : à partir de l'instant où les entrées d'un bloc combinatoire sont fixées, les sorties peuvent changer plusieurs fois avant d'adopter leurs valeurs définitives. La *logique séquentielle synchrone* laisse justement le temps aux blocs combinatoires de se stabiliser. Ce temps est en gros la période de l'horloge, en tout cas une part importante de celle-ci. C'est seulement une fois ce temps écoulé, au top d'horloge, que les bascules D viennent prendre en compte les valeurs des sorties combinatoires.

- 2d) Modifier le montage rebouclé initial pour que les valeurs de l'entrée x soient prises en compte seulement aux tops de l'horloge CK, devenant ainsi un *circuit séquentiel synchrone*.

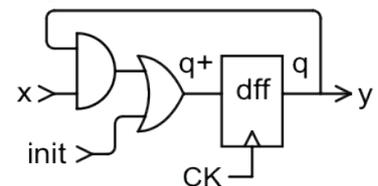
Pour cela, il faut (et il suffit) d'insérer une bascule D dans la boucle, comme montré ci-contre. L'entrée x peut désormais faire n'importe quoi entre 2 tops d'horloge (contrairement à la situation asynchrone). En revanche, elle doit être stable (à 0 ou à 1) à proximité de chaque top d'horloge. Sinon c'est une valeur inconnue qui sera prise en compte (avec certains effets secondaires possibles), sans que cela ne soit forcément dramatique au niveau du système, toutefois.



Attention : une autre possibilité aurait été d'introduire la bascule dans la partie haute de la boucle. Mais alors la sortie y aurait été de type Mealy, ce qui est interdit !

- 2e) Rajouter une fonctionnalité d'initialisation *synchrone* (c'est-à-dire prise en compte au top d'horloge seulement), commandé par *init*, qui fasse passer y à 1 quelle que soit la valeur de x .

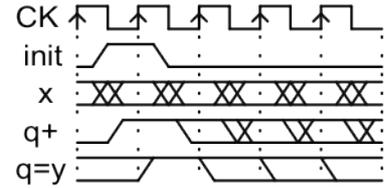
On peut insérer un MUX entre porte AND et bascule, qui laisse entrer un 1 si $init=1$ (et laisse la liaison habituelle sinon). Un tel MUX est le BDD d'une porte OU, par laquelle il peut donc être avantageusement remplacé. D'où le montage ci-contre, que l'on peut intuitivement transformer en une série de 2 NAND, dans un souci de lisibilité...



- 2f) Dessiner un chronogramme des opérations.

Selon un usage bien établi, la sortie d'une bascule D est notée q et son entrée q^+ (c'est la prochaine valeur de q). Ici $y=q$, cas simple. Les tops d'horloge sont les fronts montants de CK, tracés en pointillés de haut en bas. On commence par initialiser le montage avec *init* qui passe à 1 dès la première période d'horloge, provoquant la transition montante de q^+ peu après, et

celle de $y=q$ dès le top d'horloge suivant. Ensuite, *init* repasse à 0 et notre « détecteur synchrone de microcoupures » se retrouve armé pour détecter toute valeur de x à 0 à un top d'horloge. L'entrée x a le droit de bouger entre 2 tops d'horloge, d'où les « croisillons » qui représentent des transitions arbitraires, et en provoquent donc aussi sur q^+ . En revanche, aux tops d'horloge, x est censé valoir soit 0 soit 1 : les 2 possibilités sont représentées. Un 0 provoque la baisse définitive (en attendant une future réinitialisation) de $q=y$ dans la foulée du top d'horloge.



La logique séquentielle synchrone permet aussi de reboucler des montages combinatoires qui ne présenteraient aucun état stable binaire si rebouclés directement, tel qu'un inverseur.

2g) Décrire le comportement d'un inverseur rebouclé par une bascule D.

Sa sortie change de valeur à l'issue de chaque top d'horloge. Il s'agit donc d'un signal de fréquence 2 fois plus faible que celle de l'horloge. Ainsi, on appelle ce montage un « diviseur de fréquence par 2 ».

2h) Sur cette base, resongeant à l'oscillateur en anneau de l'exercice précédent, comment obtenir un signal qui oscillerait à quelques Hz ?

Le sortie d'un diviseur de fréquence par 2 peut être utilisé comme horloge par un autre circuit du même type... Il suffit donc d'en chaîner ainsi autant que nécessaire. Sans rien enlever à l'élégance de cette solution, soulignons qu'il est généralement très dangereux de « jouer » avec l'horloge, en s'en servant pour faire autre chose, en la combinant avec d'autres signaux ou en la remplaçant par autre chose. Hors cas très particulier comme ici, on s'interdira de le faire. En logique séquentielle synchrone, il y a une et une seule horloge et elle ne sert qu'à cadencer les bascules D. Elle n'est d'ailleurs pas considérée comme une entrée.