A2023 – INFO MPI I

ÉCOLE DES PONTS PARISTECH,
ISAE-SUPAERO, ENSTA PARIS,
TÉLÉCOM PARIS, MINES PARIS,
MINES SAINT-ÉTIENNE, MINES NANCY,
IMT ATLANTIQUE, ENSAE PARIS,
CHIMIE PARISTECH - PSL.

Concours Mines-Télécom

CONCOURS 2023

PREMIÈRE ÉPREUVE D'INFORMATIQUE

Durée de l'épreuve : 3 heures

L'usage de la calculatrice et de tout dispositif électronique est interdit.

Les candidats sont priés de mentionner de façon apparente
sur la première page de la copie :

INFORMATIQUE I - MPI

L'énoncé de cette épreuve comporte 9 pages de texte.

Cette épreuve concerne uniquement les candidats de la filière MPI.

Si, au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé,
il le signale sur sa copie et poursuit sa composition en expliquant les raisons
des initiatives qu'il est amené à prendre.

Préliminaires

Présentation du sujet

L'épreuve est composée d'un problème unique comportant 32 questions divisées en trois sections. L'objectif du problème est de construire des *listes à accès direct* : une liste à accès direct est un type de donnée abstrait qui permet, d'une part, d'empiler et de dépiler efficacement un élément en tête de liste et, d'autre part, d'accéder efficacement au k^{e} élément de la liste, pour n'importe quel indice k valide.

Dans la première section (page 1), nous étudions un système de numération : la représentation binaire gauche des entiers naturels. Dans la deuxième section (page 4), nous étudions les arbres binaires parfaits. Dans la troisième section (page 6), nous implémentons le type liste à accès direct par la structure de données concrète *liste gauche*, que nous construisons en exploitant les résultats obtenus dans les sections précédentes.

Dans tout l'énoncé, un même identificateur écrit dans deux polices de caractère différentes désignera la même entité, mais du point de vue mathématique avec la police en italique (par exemple n) et du point de vue informatique avec celle en romain avec espacement fixe (par exemple `n`).

Travail attendu

Pour répondre à une question, il est permis de réutiliser le résultat d'une question antérieure, même sans avoir réussi à établir ce résultat. Des rappels de programmation sont faits en annexe et peuvent être utilisés directement.

Selon les questions, il faudra coder des fonctions à l'aide du langage de programmation C exclusivement, en reprenant le prototype de fonction fourni par le sujet, ou en pseudo-code (c-à-d. dans une syntaxe souple mais conforme aux possibilités offertes par le langage C). Il est inutile de rappeler que les entêtes `<assert.h>`, `<stdbool.h>`, etc. doivent être inclus.

Quand l'énoncé demande de coder une fonction, sauf indication explicite de l'énoncé, il n'est pas nécessaire de justifier que celle-ci est correcte ou de tester que des préconditions sont satisfaites. On suppose que le type `int` n'est jamais sujet à des débordements.

Le barème tient compte de la clarté des programmes : nous recommandons de choisir des noms de variables intelligibles ou encore de structurer de longs codes par des blocs ou par des fonctions auxiliaires dont on décrit le rôle. Lorsqu'une réponse en pseudo-code est permise, seule la logique de programmation est évaluée, même dans le cas où un code en C a été fourni en guise de réponse.

1. Représentation binaire gauche des entiers naturels

1.1. Mise en place

Soient m un entier naturel et N un entier naturel non nul. Il est classique d'appeler *représentation binaire standard de l'entier m sur N chiffres*, ou plus simplement *représentation standard*, toute suite finie $b = (b_n)_{0 \leqslant n < N}$ de longueur N telle que, pour tout indice n compris entre 0 et $N-1$, le chiffre b_n appartient à $\{0, 1\}$ et l'égalité suivante est vérifiée

$$m = \sum_{n=0}^{N-1} b_n 2^n.$$

Définition : Nous appelons *représentation binaire gauche de l'entier m sur N chiffres* toute suite finie $g = (g_n)_{0 \leqslant n < N}$ de longueur N telle que,

(i) pour tout indice n compris entre 0 et $N-1$, le chiffre g_n appartient à $\{0, 1, 2\}$,

(ii) l'égalité suivante est satisfaite

$$m = \sum_{n=0}^{N-1} g_n(2^{n+1} - 1),$$

(iii) le chiffre « 2 » n'apparaît qu'au plus une fois parmi les chiffres $(g_n)_{0 \leqslant n < N}$,

(iv) s'il existe une position p tel que le chiffre g_p est 2, alors, pour tout indice n compris entre 0 et $p-1$, le chiffre g_n est nul.

De manière plus courte, nous parlons simplement de *représentation gauche*.

La figure 1 ci-dessous donne une représentation standard et une représentation gauche sur quatre chiffres des seize premiers entiers. Conformément à l'usage habituel, nous écrivons toute représentation, qu'elle soit standard ou gauche, sous la forme d'un mot $b_{N-1} \cdots b_0$ ou $g_{N-1} \cdots g_0$ dans lequel les chiffres de poids faibles sont écrits à droite.

Entier	Repr. standard	Repr. gauche
0	0000	0000
1	0001	0001
2	0010	0002
3	0011	0010
4	0100	0011
5	0101	0012
6	0110	0020
7	0111	0100

Entier	Repr. standard	Repr. gauche
8	1000	0101
9	1001	0102
10	1010	0110
11	1011	0111
12	1100	0112
13	1101	0120
14	1110	0200
15	1111	1000

FIGURE 1 – Représentations standard et gauche des seize premiers entiers

☐ 1 – Soit c un entier. Donner la représentation standard de l'entier dont une représentation gauche est $10 \cdots 0$ (avec c chiffres nuls) en justifiant sommairement. Faire de même avec l'entier dont une représentation gauche est $20 \cdots 0$ (avec c chiffres nuls).

☐ 2 – Déterminer, en justifiant, le plus grand entier naturel M_N qui admet une représentation gauche sur N chiffres. Préciser la représentation gauche de cet entier.

Définition : Soit n_0 un indice compris entre 0 et $N-1$. On dit que l'indice n_0 est la *position du chiffre de plus fort poids* d'une représentation $g_{N-1} \cdots g_0$ si l'indice n_0 est le plus petit entier tel que, pour tout indice $n > n_0$, le chiffre g_n est nul. On appelle le chiffre g_{n_0} le *chiffre de plus fort poids*.

☐ 3 – Soient N un entier naturel non nul, $g = g_{N-1} \cdots g_0$ et $h = h_{N-1} \cdots h_0$ deux représentations gauches d'un même entier m. Démontrer que les chiffres de plus fort poids de g et de h sont de même valeur et à la même position.

☐ 4 – Démontrer que tout entier appartenant à l'intervalle $[\![0, M_N]\!]$, où l'entier M_N a été introduit à la question 2, ne possède au plus qu'une seule représentation gauche sur N chiffres.

Indication C : L'entier N est déclaré comme constante globale. Nous utilisons la structure C déclarée comme suit pour écrire la représentation gauche sur N chiffres d'un entier $g_{N-1} \cdots g_0$:

```
1.  const int N = 8;
2.
3.  struct RepGauche {
4.    int position;
5.    bool chiffres[N];
6.  };
7.  typedef struct RepGauche rg;
```

Le champ `position` repère la position éventuelle du chiffre 2 ; il vaut -1 au cas où le chiffre 2 n'apparaît pas. Pour tout indice n compris entre 0 et $N-1$, la n^{e} case du champ `chiffres` vaut `true` si le chiffre g_n est non-nul et vaut `false` sinon.

Par exemple, les entiers 15 et 21 ont pour représentation gauche les variables `entier_15` et `entier_21` suivantes :

```
8.  rg entier_15 = { .position = -1,
9.                   .chiffres = { 0, 0, 0, 1, 0, 0, 0, 0 } };
10. rg entier_21 = { .position =  1,
11.                  .chiffres = { 0, 1, 0, 1, 0, 0, 0, 0 } };
```

☐ 5 – Formaliser sous la forme d'un ou de plusieurs invariants le fait qu'une valeur C de type `rg` est la représentation gauche d'un entier.

☐ 6 – Écrire une fonction C `int rg_to_int(rg g)`, qui renvoie l'entier dont g est la représentation gauche. On supposera que l'invariant de la question 5 est satisfait.

1.2. Incrémentation et décrémentation

Nous proposons l'algorithme suivant :

ALGORITHME MYSTÈRE :
Entrée : Représentation gauche $g = g_{N-1} \cdots g_0$ d'un certain entier m.
Effet :

- Si aucun des chiffres $(g_n)_{0 \leqslant n < N}$ ne vaut 2, changer le chiffre g_0 en $g_0 + 1$.
- Sinon, en notant p la position du chiffre 2, changer le chiffre g_p en 0 et le chiffre g_{p+1} en $g_{p+1} + 1$.

Nous notons m' l'entier dont la représentation gauche est g après exécution de l'algorithme.

FIGURE 2 – Un algorithme

☐ 7 – Vérifier que l'invariant de la question 5 n'est pas rompu par l'algorithme mystère (*cf.* figure 2). Avec les notations m et m' de la figure 2, caractériser, en fonction de l'entier m, la valeur de l'entier m'.

☐ 8 – Écrire une fonction C `bool rg_incr(rg *s)` dont la spécification suit :
Précondition : La variable s est un pointeur vers la représentation gauche d'un entier m.
Effet : La valeur pointée par s est modifiée afin de représenter l'entier $m+1$.
Valeur de retour : Booléen `true` si l'incrémentation de m peut avoir lieu et `false` si un débordement se produit car $m+1$ n'est pas représentable sur le même nombre de chiffres.

☐ 9 – Calculer la complexité en temps dans le pire des cas de la fonction `rg_incr` en fonction de N. Comparer avec la complexité de la même opération sur la représentation standard.

☐ 10 – Écrire une fonction C `bool rg_decr(rg *s)` dont la spécification suit :
Précondition : Le pointeur s désigne la représentation gauche d'un entier m.
Effet : La valeur pointée par s est modifiée afin de représenter l'entier $m-1$.
Valeur de retour : Booléen `true` si la décrémentation de m peut avoir lieu et `false` si un débordement se produit car m est nul.
Il est recommandé d'expliquer son intention avant de donner son code.

2. Arbres binaires parfaits

2.1. Opérations sur les arbres binaires

Indication C : Nous représentons des arbres binaires à valeurs entières au moyen du type C `arb` suivant, qui est un pointeur vers une structure contenant la valeur entière dans le champ `valeur` et les deux fils dans les champs `fils_g` et `fils_d`. L'arbre vide se représente par le pointeur `NULL`.

```
12.  typedef struct Noeud *arb;
13.  struct Noeud {
14.    int valeur;
15.    arb fils_g;
16.    arb fils_d;
17.  };
```

L'arbre vide est, par convention, de hauteur -1.

☐ 11 – Écrire une fonction C `int hauteur(arb a)` qui calcule la hauteur de l'arbre a.

☐ 12 – Écrire une fonction C `arb noeud(int v, arb ag, arb ad)` qui construit un arbre dont la racine a pour étiquette v, dont le fils gauche est a_g et le fils droit est a_d. Dans cette question, il est demandé de se défendre, par le truchement d'une assertion, de toute erreur liée à un échec d'allocation dynamique de mémoire.

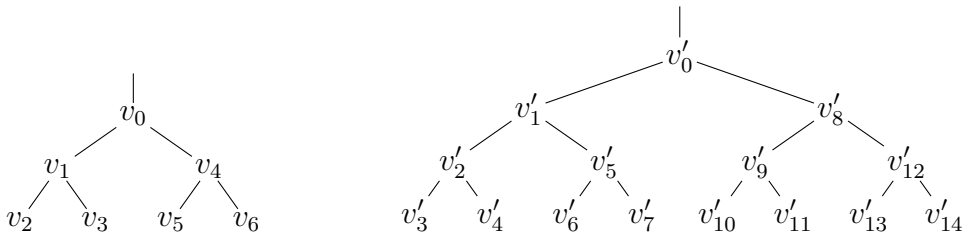

FIGURE 3 – Arbres binaires parfaits de hauteur 2 et de hauteur 3.

2.2. Arbres parfaits

Définition : Un *arbre binaire parfait*, ou simplement *arbre parfait*, est un arbre binaire dont tous les nœuds internes ont exactement deux fils, dont toutes les feuilles sont à la même profondeur et dont les nœuds sont étiquetés par des valeurs entières.

Des exemples d'arbres binaires parfaits sont donnés figure 3.

☐ 13 – Démontrer que tout arbre binaire parfait de hauteur n possède $2^{n+1} - 1$ nœuds.

☐ 14 – Écrire une fonction C `bool est_parfait(arb a, int n)` dont la spécification suit :
Précondition : Le pointeur a désigne la racine d'un arbre binaire (dont les nœuds sont bien tous distincts).
Valeur de retour : Booléen `true` si l'arbre pointé est parfait de hauteur n et `false` sinon.

☐ 15 – Calculer la complexité en temps dans le pire des cas de l'exécution de `est_parfait(a, n)` en fonction de l'entier n.

2.3. Opérations sur les arbres parfaits

Définition : Soient S une structure de données qui permet de stocker une collection d'entiers et t le cardinal de S. Nous disons que la structure de données S est à *accès direct* s'il existe une manière systématique de numéroter chaque élément de S entre 0 et $t - 1$ et s'il existe deux primitives, `acces` et `modif`, qui permettent respectivement de consulter et de modifier n'importe quel élément de S en temps logarithmique en t à partir seulement de son numéro.

Nous souhaitons vérifier qu'un arbre parfait est une structure de données à accès direct. Nous numérotons les éléments d'un arbre parfait en utilisant l'ordre préfixe de gauche à droite de l'arbre (comme illustré dans la figure 3).

☐ 16 – Écrire une fonction C `arb arb_trouve(arb a, int n, int k)` ainsi spécifiée :
Précondition : Le pointeur a désigne la racine d'un arbre binaire parfait de hauteur n. L'entier k satisfait les inégalités $0 \leqslant k < 2^{n+1} - 1$.
Valeur de retour : Pointeur vers le k^{e} nœud de l'arbre dans l'ordre préfixe.

Il est rappelé que la racine d'un arbre est à profondeur 0. Nous donnons la formule de sommation, valable pour tout réel $a \neq 1$ et tout entier t,

$$\sum_{k=0}^{t} k \cdot a^k = \frac{((a-1)t - 1)\, a^{t+1} + a}{(a-1)^2}.$$

☐ 17 – Nous appelons P la profondeur d'un nœud choisi aléatoirement et uniformément parmi les $2^{n+1}-1$ nœuds d'un arbre binaire parfait de hauteur n. Calculer l'espérance $\mathbb{E}(P)$ de la variable aléatoire P.

☐ 18 – Déterminer la complexité moyenne en temps de l'instruction `arb_trouve(a, n, k)` lorsque la hauteur n et l'arbre a sont fixés et le numéro k est choisi aléatoirement et uniformément dans l'intervalle $[\![0, 2^{n+1}-2]\!]$.

☐ 19 – Écrire une fonction C `int arb_acces(arb a, int n, int k)` qui renvoie la k^{e} valeur de l'arbre a de hauteur n ainsi qu'une fonction C `void arb_modif(arb a, int n, int k, int v)` qui remplace la k^{e} valeur de l'arbre a de hauteur n par la valeur v. Dire finalement si la structure de données arbre binaire parfait est à accès direct.

3. Listes gauches

Les arbres binaires parfaits seuls ne permettent de représenter que des collections d'entiers dont le cardinal est de la forme $2^{n+1}-1$ où n est entier naturel. Afin de représenter des collections de cardinal quelconque, nous introduisons des suites d'arbres parfaits, aux hauteurs strictement croissantes et savamment choisies.

Définition : Nous appelons *liste binaire gauche de cardinal m sur N arbres* la structure de données constituée de $N+1$ arbres binaires parfaits $((a_n)_{0\leqslant n<N}, e)$ comme suit. Nous décomposons l'entier m selon sa représentation binaire gauche sur N chiffres : $g_{N-1}\dots g_0$. Pour tout indice n compris entre 0 et $N-1$, si le chiffre g_n n'est pas nul, l'arbre binaire parfait a_n est un arbre de hauteur n, sinon il s'agit de l'arbre vide. Si le chiffre 2 apparaît parmi les chiffres de la représentation gauche de m et si l'indice p est sa position, alors l'arbre e est un arbre binaire parfait de hauteur p, sinon l'arbre e est l'arbre vide. De manière plus courte, nous parlons simplement de *liste gauche*.

Une liste binaire gauche de cardinal m sur N arbres permet de stocker une collection de m éléments entiers. Il apparaît qu'avec N arbres, une liste binaire gauche peut stocker jusqu'à M_N entiers (où l'entier M_N a été introduit à la question 2) et que m est inférieur à M_N.

Indication C : Nous adoptons le type C suivant

```
18.  struct ListeGauche {
19.    int hauteur_e;
20.    arb extra;
21.    int nb_arbres;
22.    arb *arbres;
23.  };
24.  typedef struct ListeGauche lg;
```

Le champ `hauteur_e` contient la hauteur p de l'arbre exceptionnel e. Le champ `extra` désigne la racine de l'arbre exceptionnel e. Le champ `nb_arbres` contient l'entier N. Enfin, le champ `arbres` désigne un tableau de N pointeurs vers les arbres $(a_n)_{0\leqslant n<N}$ qui a été alloué dynamiquement.

3.1. Opérations simples sur les listes gauches

☐ 20 – Écrire une fonction C `lg lg_init(int N)` qui renvoie une liste gauche de cardinal nul sur N arbres.

☐ 21 – Écrire une fonction C `int lg_card(lg l)` qui calcule le cardinal m de la liste gauche ℓ.

Définition : Afin de numéroter l'ensemble des éléments d'une liste gauche $\ell = ((a_n)_{0 \leqslant n < N}, e)$, nous parcourons d'abord les éléments de l'arbre exceptionnel e dans l'ordre préfixe, puis nous parcourons les éléments des arbres a_0, a_1, …, a_{N-1} dans l'ordre préfixe. Les numéros sont attribués aux valeurs rencontrées par ordre de première rencontre.

☐ 22 – Écrire une fonction C `arb lg_trouve(lg l, int k)` qui renvoie le k^{e} nœud de la liste gauche ℓ. On supposera que k est un indice valide ($0 \leqslant k < m$).

☐ 23 – Calculer la complexité en temps dans le pire des cas de `lg_trouve l` en fonction de la capacité maximale M_N de la liste gauche ℓ.

☐ 24 – Écrire une fonction C `int lg_acces(lg l, int k)` qui renvoie la k^{e} valeur de la liste gauche ℓ ainsi qu'une fonction C `void lg_modif(lg l, int k, int v)` qui remplace la k^{e} valeur de la liste gauche ℓ par la valeur v.

3.2. Ajout et suppression en tête de liste gauche

☐ 25 – Soient v une valeur entière et $\ell = ((a_n)_{0 \leqslant n < N}, e)$ une liste gauche de cardinal m, avec $m < M_N$, qui contient les éléments v_1, …, v_m dans cet ordre. Décrire, en fonction de la liste gauche ℓ, la liste gauche $\ell' = ((a'_n)_{0 \leqslant n < N}, e')$ de cardinal $m+1$ dont les éléments sont v, v_1, …, v_m dans cet ordre. En déduire le principe d'une fonction C `bool lg_empile(int v, lg l)` réalisant l'insertion de la valeur v en tête de la liste gauche ℓ où le booléen résultat vaut `true` si l'ajout a eu lieu et vaut `false` si un débordement de capacité de la liste se produit. On ne demande pas le code complet de cette fonction.

☐ 26 – Déterminer la complexité en temps dans le pire des cas de la fonction `lg_empile`.

☐ 27 – Donner le principe d'une fonction C `bool lg_depile(int *w, lg l)` réalisant le retrait de l'élément de tête de la liste gauche ℓ et son affectation à l'adresse w. Le booléen résultat vaut `true` si le retrait a eu lieu et vaut `false` si l'opération a échoué en raison d'une liste vide. On ne demande pas le code complet.

☐ 28 – Donner la complexité en temps dans le pire des cas de la fonction `lg_depile`.

☐ 29 – Discuter la possibilité d'obtenir une complexité plus faible à la question 28, quitte à modifier légèrement la définition du type `lg`.

☐ 30 – Soit N un entier et $\ell = ((a_n)_{0 \leqslant n < N}, e)$ une liste gauche. Discuter la possibilité de modifier en place et avec une faible complexité en temps la liste gauche ℓ de sorte que le nombre d'arbres N devienne $N + 1$ et que les mêmes éléments demeurent dans la liste.

3.3. Utilisation concurrente des listes gauches

Dans cette sous-section, on raisonne sur les algorithmes en supposant que les fils d'exécution peuvent s'entrelacer mais que les instructions d'un même fil s'exécutent dans l'ordre du programme. L'entête `#include <pthread.h>` a été déclarée ; la syntaxe de certaines fonctions s'y rattachant est rappelée en annexe.

☐ 31 – Deux fils d'exécution distincts exécutent la fonction `lg_empile` sur la même liste gauche. Montrer qu'une course critique advient de leur exécution concurrente et que la cohérence de ladite liste gauche n'est pas garantie, autrement dit que certains invariants qui caractérisent la bonne formation d'une liste gauche peuvent être violés à l'issue des exécutions.

Nous nous intéressons finalement au *problème des producteurs et des consommateurs* qui s'échangent des entiers au travers d'un tampon, ici constitué par une unique liste gauche ℓ à N arbres, l'entier N étant fixé à l'avance. Les producteurs écrivent dans la liste gauche ℓ en empilant un entier à la fois en tête, à condition que la liste gauche ℓ ne soit pas pleine ; les consommateurs vident la liste gauche ℓ en dépilant l'entier en tête de la liste, à condition que la liste gauche ℓ ne soit pas vide.

Un seul agent peut accéder au tampon à la fois. Lorsqu'un consommateur souhaite supprimer une donnée alors que le tampon est vide, il est mis en attente ; lorsqu'un producteur souhaite écrire une donnée alors que le tampon est plein, il est mis en attente.

☐ 32 – Décrire une solution au problème des producteurs et des consommateurs en s'appuyant sur un verrou et sur deux sémaphores. Détailler sous la forme de code C ou bien de pseudo-code ce que font les producteurs et les consommateurs.

* *
*

Les listes binaires gauches, ou *skew binary lists*, ont été inventées par Eugene Myers en 1983.

A. Rappels de programmation en C

L'expression `1 << n` représente le décalage de la valeur 1 sur n bits : elle a pour valeur l'entier 2^n.

Le type **`pthread_t`** désigne des fils d'exécution.

L'instruction `pthread_create(`**`pthread_t`** `*th_id, NULL, &ma_fonction,` **`void`** `*args)` crée un nouveau fil d'exécution qui appelle la fonction `ma_fonction` sur le ou les arguments désignés par `args` et qui s'exécute simultanément avec le fil d'exécution appelant.

L'instruction `pthread_join(th_id, NULL)` suspend l'exécution du fil d'exécution appelant jusqu'à ce que le fil d'exécution identifié par `th_id` achève son exécution.

Le type **`pthread_mutex_t`** désigne des verrous.

L'instruction `pthread_mutex_lock(&v)` verrouille le verrou `v`.

L'instruction `pthread_mutex_unlock(&v)` déverrouille le verrou `v`.

Le type **`sem_t`** désigne des sémaphores.

L'instruction `sem_init(&s, 0, v)` initialise le sémaphore `s` à la valeur v (avec $v \geqslant 0$).

L'instruction `sem_wait(&s)` décrémente le compteur du sémaphore `s` : si le compteur est toujours positif, l'appel se termine ; sinon le fil d'exécution appelant est bloqué.

L'instruction `sem_post(&s)` incrémente le compteur du sémaphore `s` et, si le compteur redevient strictement positif, réveille un fil d'exécution bloqué sur `s`.

Fin de l'épreuve