

# ASYMPTOTIC BOUNDS FOR SEPARATING SYSTOLES ON SURFACES

STÉPHANE SABOURAU

ABSTRACT. The separating systole on a closed Riemannian surface  $M$ , denoted by  $\text{sys}_0(M)$ , is defined as the length of the shortest noncontractible loops which are homologically trivial. We answer positively a question of M. Gromov [Gr96, 2.C.2.(d)] about the asymptotic estimate on the separating systole. Specifically, we show that the separating systole of a closed Riemannian surface  $M$  of genus and area  $g$  satisfies an upper bound similar to M. Gromov's asymptotic estimate on the (homotopy) systole. That is,  $\text{sys}_0(M) \lesssim \log g$ .

## CONTENTS

1. Introduction	1
2. Systolic inequalities on graphs	4
3. Morse functions and first Betti numbers	7
4. Construction of graphs on the surface	9
5. Fundamental groups of graphs and surfaces	12
6. Proof of the main theorem	17
References	20

## 1. INTRODUCTION

Let  $M$  be a nonsimply connected closed Riemannian surface. The (homotopy) systole of  $M$ , denoted by  $\text{sys } \pi_1(M)$  or  $\text{sys}(M)$  for short, is defined as the length of the shortest noncontractible loops in  $M$ .

We define the optimal systolic area of a nonsimply connected closed surface  $M$  as

$$\sigma(M) = \inf \frac{\text{Area}(M)}{\text{sys}(M)^2} \quad (1.1)$$

---

2000 *Mathematics Subject Classification.* Primary 53C22; Secondary 05C38.

*Key words and phrases.* systole, separating systole, systolic area, systolic ratio, asymptotic bound, surface, graph.

where the infimum is taken over the space of all the metrics on  $M$ . The optimal systolic area is a topological invariant of surfaces.

The exact value of the optimal systolic area is known for the torus, *cf.* [Be93], the projective plane [Pu52] and the Klein bottle [Ba86]. For a notion of systole extended to the isometry groups of Riemannian manifolds, the optimal systolic area has also been computed for the 17 crystallographic groups of the plane and the triangle groups [Ba93]. No other exact value of the optimal systolic area is known.

However, nontrivial lower bounds on the optimal systolic area of every nonsimply connected closed surface have been established, *cf.* [Gr83], [Gr96], [KS06a], [KS05], [KS06b] and [Sa06a] for recent developments. For instance, we deduce from [Pu52] and [Gr83, 5.2.B] that every nonsimply connected closed surface  $M$  satisfies

$$\sigma(M) \geq \frac{2}{\pi} \tag{1.2}$$

with equality if and only if  $M$  is homeomorphic to the projective plane. Such an inequality is called a systolic inequality.

In higher dimension, nonoptimal systolic inequalities exist for essential manifolds [Gr83] and optimal systolic inequalities exist for the notions of one-dimensional stable and conformal systoles, *cf.* [Gr99], [IK04], [BCIK]. We refer to the expository texts [Be93], [Gr96], [Gr99] and [CK03], and the references therein for an account on the subject and other generalizations in higher dimensions.

The systolic inequality (1.2) can be improved by taking into account the topology of  $M$ . For instance, the following result of M. Gromov shows that closed surfaces of large genus have a large optimal systolic area.

**Theorem 1.1** ([Gr83, 6.4.D'], [Gr96, 3.C.3]). *There exists a positive constant  $C$  such that every closed surface  $M$  of genus  $g$  satisfies*

$$\sigma(M) \geq C \frac{g}{(\ln g)^2}. \tag{1.3}$$

It is shown in [KS05], using different techniques, that (1.3) holds for every  $C < \pi$  provided that  $g$  is large enough. We refer to the end of this section for a discussion about the different proofs of this result. See also [Gr83], [Gr96] and [Sa06b] for generalizations in higher dimensions.

An upper bound on the optimal systolic area of surfaces of large genus has been found by P. Buser and P. Sarnak [BS94]. Namely, they construct hyperbolic surfaces  $(M, hyp)$  of genus  $g$ , obtained as

coverings of an arithmetic Riemann surface, such that

$$\text{sys}(M, \text{hyp}) \gtrsim \ln g \quad (1.4)$$

Other constructions and higher dimension generalizations can be found in [KSV].

Combining the inequalities (1.3) and (1.4), we obtain

$$C \frac{g}{(\ln g)^2} \leq \sigma(M) \leq C' \frac{g}{(\ln g)^2} \quad (1.5)$$

where  $C$  and  $C'$  are two universal positive constants.

In particular, there is no lower bound on  $\sigma(M)$  linear in the genus  $g$  (see [BB05] for a generalization in higher dimension).

Given a closed Riemannian surface  $M$  of genus  $g \geq 2$ , we define the separating systole, denoted by  $\text{sys}_0(M)$ , as the length of the shortest noncontractible loops in  $M$  which are homologically trivial. That is,  $\text{sys}_0(M) = \inf\{\text{length}(\gamma) \mid \gamma \text{ induces a nontrivial class in } [\pi_1(M), \pi_1(M)]\}$ .

In this definition, the homology coefficients are in  $\mathbb{Z}$  and the commutator of the fundamental group of  $M$  is noted  $[\pi_1(M), \pi_1(M)]$ .

Similar to the optimal systolic area  $\sigma(M)$ , we define

$$\sigma_0(M) = \inf \frac{\text{Area}(M)}{\text{sys}_0(M)^2}$$

where the infimum is taken over the space of all the metrics on  $M$ .

We clearly have  $\text{sys}(M) \leq \text{sys}_0(M)$ , *cf.* (1.8) for a more precise estimate, hence  $\sigma(M) \geq \sigma_0(M)$ .

Lower bounds on  $\sigma_0(M)$  can be deduced from [Gr83, 5.4] (indeed, the length of the commutator of two independent loops gives an upper bound on the separating systole). In particular, for surfaces of large genus, M. Gromov showed the following.

**Theorem 1.2** ([Gr83, 5.4.B], [Gr96, 2.C.2.(d)]). *For every  $\alpha < 1$ , there exists a positive constant  $C_\alpha$  such that every closed surface  $M$  of genus  $g \geq 2$  satisfies*

$$\sigma_0(M) \geq C_\alpha g^\alpha.$$

The hyperbolic surfaces  $(M, \text{hyp})$  constructed in [BS94] satisfy

$$\text{sys}_0(M, \text{hyp}) \gtrsim \ln g. \quad (1.6)$$

Therefore, Theorem 1.2 does not hold for  $\alpha = 1$ .

M. Gromov asked in [Gr96, 2.C.2.(d)] whether a lower bound on  $\sigma_0(M)$  similar to (1.3) exists. We answer this question positively. Specifically, we prove the following.

**Main Theorem 1.3.** *There exists a positive constant  $C$  such that every closed surface  $M$  of genus  $g \geq 2$  satisfies*

$$\sigma_0(M) \geq C \frac{g}{(\ln g)^2}. \quad (1.7)$$

The estimate (1.6) shows that the inequality (1.7) yields the right asymptotic bound. More precisely, a double inequality similar to (1.5) holds for  $\sigma_0(M)$ .

Such a double inequality also holds for  $\sigma_H(M)$ , where  $\sigma_H(M)$  is defined by replacing in (1.1) the systole with the homology systole, *cf.* [Gr96]. Recall that the homology systole, denoted by  $\text{sys}_H(M)$ , is the length of the shortest homologically nontrivial loops in  $M$ .

We clearly have  $\text{sys}(M) \leq \text{sys}_H(M)$  but no lower bound on  $\sigma_0(M)$  can immediately be deduced from a lower bound on  $\sigma_H(M)$ .

Note also that

$$\text{sys}(M) = \min\{\text{sys}_0(M), \text{sys}_H(M)\}. \quad (1.8)$$

The systolic inequality (1.3) was first proved in [Gr83, p. 74] by using a technique known as "diffusion of chains". It was then improved in [KS05], where an upper bound for the entropy of a systolically extremal surface in terms of its systole is established. As observed by F. Balacheff [B04], the inequality (1.3) can also be derived by combining the works of S. Kodani [Ko87], and B. Bollobás, E. Szemerédi and A. Thomason [BT97, BS02] on systolic inequalities of graphs.

We will follow this latter approach to prove Main Theorem 1.3. More precisely, we construct a graph on a surface (Section 4), establish systolic inequalities on this graph (Section 2), relate the length of cycles on the graph to the separating systole on the surface (Section 5) and deduce a lower bound on the area of the surface through a core formula as in [Ko87] (Section 6).

*Acknowledgment.* The author would like to thank the referee for his or her comments, especially on systolic inequalities on graphs, *cf.* Remark 2.3.

## 2. SYSTOLIC INEQUALITIES ON GRAPHS

By definition, a metric graph is a graph (*i.e.*, a finite 1-dimensional simplicial complex) endowed with a length structure.

The homotopy class of a graph  $\Gamma$  is characterized by its first Betti number  $b(\Gamma)$ , which can be computed as follows:

$$b(\Gamma) = e(\Gamma) - v(\Gamma) + n(\Gamma) \quad (2.1)$$

where  $e(\Gamma)$ ,  $v(\Gamma)$  and  $n(\Gamma)$  are respectively the number of edges, vertices and connected components of  $\Gamma$ .

**Definition 2.1.** The systole and the total length of a metric graph  $\Gamma$ , denoted by  $\text{sys}(\Gamma)$  and  $\text{length}(\Gamma)$ , are respectively defined as the length of the shortest noncontractible loop of  $\Gamma$  and the sum of the lengths of the edges of  $\Gamma$ . These two metric invariants do not depend on the simplicial structure of the graph.

We define the systolic length of  $\Gamma$  as

$$\sigma'(\Gamma) = \frac{\text{length } \Gamma}{\text{sys}(\Gamma)}.$$

For every  $b \geq 1$ , we also define a homotopy invariant of a graph as

$$\sigma'(b) = \inf \{ \sigma'(\Gamma) \mid \Gamma \text{ metric graph with first Betti number } b \}.$$

We put a ' when we deal with graphs (as in  $\sigma'(\Gamma)$  or  $\sigma'(b)$ ) and no ' when we deal with surfaces (as in  $\sigma(M)$ ).

In [BS02], B. Bollobàs and E. Szemerédi improved the multiplicative constant of a systolic inequality for graphs established by B. Bollobàs and A. Thomasson [BT97]. Specifically, they proved

**Theorem 2.2** ([BS02]). *For every  $b \geq 3$ , we have*

$$\sigma'(b) \geq \frac{3}{2} \frac{b-1}{\log_2(b-1) + \log_2 \log_2(b-1) + 4} \quad (2.2)$$

where  $\log_2$  is the logarithm to the base 2.

In particular, for every  $b \geq 2$ , we have

$$\sigma'(b) \geq \frac{1}{4} \frac{b}{\ln b}. \quad (2.3)$$

Strictly speaking (2.3) is a consequence of (2.2) only for  $b \geq 5$ . When  $2 \leq b \leq 4$ , the bound (2.3) still holds because the systolic length of a graph is at least 1, that is  $\sigma'(b) \geq 1$ .

**Remark 2.3.** The referee of this article pointed out to the author that a bound similar to (2.3) (with a slightly worse constant but this does not matter for our purpose) can be obtained from elementary arguments that do not rely on the main result of [BS02]. Indeed, the estimates (1) and (2) of [BS02, §2] allow us to extend the inequality [BT97, Theorem 5] stated for non-weighted graphs to metric graphs.

Combined with the upper bound on the systolic length of graphs established in [BB05], we have, for every  $b \geq 2$ ,

$$\frac{1}{4} \frac{b}{\ln b} \leq \sigma'(b) \leq 8 \ln(2) \frac{b}{\ln b}.$$

**Definition 2.4.** A metric graph  $\Gamma$  is said to be *admissible* if the first Betti number of each of its connected components is at least 2.

As in the case of surfaces, we define the separating systole of an admissible graph  $\Gamma$  as

$$\text{sys}_0(\Gamma) = \inf_{\Lambda} \{\text{length}(\gamma) \mid \gamma \text{ induces a nontrivial class in } [\pi_1(\Lambda), \pi_1(\Lambda)]\}$$

where  $\Lambda$  runs over the connected components of  $\Gamma$ .

As previously, we also define

$$\sigma'_0(\Gamma) = \frac{\text{length } \Gamma}{\text{sys}_0(\Gamma)} \tag{2.4}$$

and, for every  $b \geq 2$ ,

$$\sigma'_0(b) = \inf \{\sigma'_0(\Gamma) \mid \Gamma \text{ admissible metric graph with first Betti number } b\}.$$

Using Theorem 2.2, we obtain a lower bound on  $\sigma'_0(b)$ . More precisely, we show

**Proposition 2.5.** *For every  $b \geq 2$ , we have*

$$\sigma'_0(b) \geq \frac{1}{64} \frac{b}{\ln b}.$$

*Proof.* Let  $\Gamma$  be an admissible metric graph whose first Betti number equals  $b$ . Without loss of generality, we can assume that  $\Gamma$  is connected and that  $\text{sys}_0(\Gamma) = 1$ . Denote by  $k$  the number of “small cycles”, *i.e.*, the number of simple loops of length less than  $\frac{1}{8}$ .

Suppose that two “small cycles”,  $c_1$  and  $c_2$ , intersect each other. The homotopy classes of these two loops (based at the same point) do not commute. Therefore, the loop  $c_1 \cup c_2 \cup c_1^{-1} \cup c_2^{-1}$ , of length

$$2 \text{length}(c_1) + 2 \text{length}(c_2) \leq \frac{1}{2},$$

which represents their commutator, is not contractible in  $\Gamma$ . Thus,  $\text{sys}_0(\Gamma) \leq \frac{1}{2}$ , hence a contradiction.

Therefore, the “small cycles” of  $\Gamma$  are disjoint.

Let  $c$  be a shortest path between two “small cycles”  $c_1$  and  $c_2$  (recall that  $\Gamma$  is connected). The commutator of the homotopy classes of  $c_1$  and  $c \cup c_2 \cup c^{-1}$  can be represented by a loop of length at most  $\frac{1}{2} + 2 \text{length}(c)$ . Thus,  $\text{sys}_0(\Gamma) \leq \frac{1}{2} + 2 \text{length}(c)$ . Therefore, the length of  $c$  is at least  $\frac{1}{4}$ .

We deduce that the  $\frac{1}{8}$ -(open)-neighborhoods of the “small cycles” are disjoint. Since  $\Gamma$  is connected, the length of the  $\frac{1}{8}$ -neighborhood of each “small cycle”  $c_i$  is at least  $\text{length}(c_i) + \frac{1}{8}$ . Therefore,  $\text{length}(\Gamma) \geq \frac{k}{8}$ .

If  $k \geq \frac{b}{2}$ , then  $\text{length}(\Gamma) \geq \frac{b}{16}$ . Thus, for every  $b \geq 2$ , we have

$$\sigma'_0(b) \geq \frac{b}{16} \geq \frac{1}{64} \frac{b}{\ln b}.$$

If  $k < \frac{b}{2}$ , we remove an edge from each “small cycle”. This gives rise to a graph  $\Gamma' \subset \Gamma$  with first Betti number  $b' = b - k$  (recall that the “small cycles” are disjoint). By construction, the systole of  $\Gamma'$  is at least  $\frac{1}{8}$ , *i.e.*,  $\text{sys}(\Gamma') \geq \frac{1}{8}$ . Further,  $\text{length}(\Gamma) \geq \text{length}(\Gamma')$ . Therefore,

$$\sigma'_0(\Gamma) \geq \frac{1}{8} \sigma'_0(\Gamma').$$

Since  $\frac{b}{2} < b' \leq b$ , the inequality (2.3) implies that

$$\sigma'_0(\Gamma) \geq \frac{1}{8} \frac{b/2}{4 \ln b}.$$

Hence the result.  $\square$

### 3. MORSE FUNCTIONS AND FIRST BETTI NUMBERS

Recall the definition of a topological Morse function. We refer to [Mo59] and [Mo75] for a general study and applications.

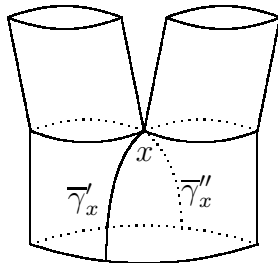
**Definition 3.1.** Let  $f$  be a continuous function on a closed  $n$ -manifold  $M$ . A point  $x$  of  $M$  is said to be *regular* (resp. *critical* of index  $p$ ) if there exists a topological chart  $\phi$  at  $x$  such that  $f \circ \phi^{-1} - f(x)$  is a linear projection (resp. the restriction of a quadratic form of signature  $(n-p, p)$ ). In these definitions, the chart  $\phi$  is not necessarily a diffeomorphism, only a homeomorphism. Furthermore, these definitions do not depend on the choice of the chart  $\phi$ .

The function  $f$  is a *topological Morse function* if every point of  $M$  is regular or critical of index  $p$  for some integer  $p$ . Note that a topological Morse function on a closed manifold has finitely many critical points.

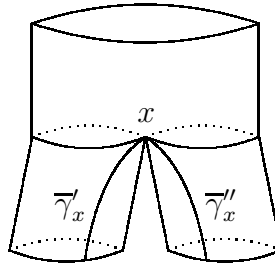
Let  $f$  be a topological Morse function on a surface  $M$  of genus  $g$  with only one critical point on each critical level. Let  $x$  be a critical point of index 1 and  $y = f(x)$ . Since  $f$  is a topological Morse function with only one critical point on each critical level, the connected component of  $f^{-1}(y)$  containing the point  $x$  is a union  $Z_x$  of two simple loops intersecting only at  $x$ .

**Definition 3.2.** We say that  $x$  is of type I if the intersection of every  $r$ -neighborhood  $U_r(Z_x)$  of  $Z_x$  with  $f^{-1}(]y-r, y[)$  has one connected component (one cylinder) for  $r$  small enough.

Otherwise, we say that  $x$  is of type II. In this case, the intersection of every  $r$ -neighborhood  $U_r(Z_x)$  of  $Z_x$  with  $f^{-1}(]y-r, y[)$  has two connected components (two cylinders) for  $r$  small enough. Recall that  $M$  is orientable,  $f$  is a topological Morse function and  $x$  is of index 1.



Type I: one leg



Type II: two legs

The following result can be found in [Ko87, (4.9)]. We include a proof for the sake of completeness.

**Lemma 3.3.** *Suppose that  $f$  has only one local minimum. Then, the function  $f$  has exactly  $g$  points of type II.*

*Proof.* Let  $N_i$  be the number of critical points of index  $i$ . From Morse formula [Mo59], [Mo75, Theorem 10.1], we have

$$1 - N_1 + N_2 = 2 - 2g. \quad (3.1)$$

We also have

$$N_1 = N_I + N_{II}, \quad (3.2)$$

where  $N_I$  and  $N_{II}$  are the numbers of critical points of type I or II. We can lower the critical points of type I (as in [Ma02, Theorem 2.34] for instance), preserving the total number of critical points, their index and their type, so that all the critical points of type I for the new function, still denoted  $f$ , lie in  $f^{-1}(]-\infty, t_0[)$  and all the critical points of type II or index 2 lie in  $f^{-1}(]t_0, \infty[)$  for some  $t_0$ . As  $t$  increases, every time we pass a critical value corresponding to a critical point of type I, the number of holes of  $f^{-1}(]-\infty, t[)$  increases by one. Therefore,  $f^{-1}(]-\infty, t_0[)$  is homeomorphic to a sphere with  $N_I + 1$  holes. On the other hand, since no critical point of type I lies in  $f^{-1}(]t_0, \infty[)$ , the space  $f^{-1}(]t_0, \infty[)$  has no handle. Further, each connected component of  $f^{-1}(]t_0, \infty[)$  has only one critical point of index 2. We conclude

as previously that  $f^{-1}(]t_0, \infty[)$  is a union of punctured spheres with  $N_{II} + N_2$  boundary components. Therefore,

$$N_I + 1 = N_{II} + N_2.$$

Combining this equation with (3.1) and (3.2), we derive the desired result, *i.e.*,  $N_{II} = g$ .  $\square$

#### 4. CONSTRUCTION OF GRAPHS ON THE SURFACE

Let  $M$  be a real analytic Riemannian surface of genus  $g \geq 2$ . Fix a point  $x_0$  lying in a systolic loop of  $M$ . Denote by  $F$  the distance function from  $x_0$ , *i.e.*,  $F(x) = \text{dist}(x_0, x)$  for every  $x \in M$ .

**Proposition 4.1.** *The function  $F$  is a topological Morse function.*

*Proof.* Every point lying outside the cut locus of  $x_0$  is regular. Therefore, we will only consider points in the cut locus of  $x_0$ .

Now the result follows from the study of the cut locus in [Be77, pp. 194-198] (see also [Fi40]), where [Be77, Lemme 3] is replaced by [He82, Lemma 1.2]. As observed in [He82], the condition of bisection and the structure of the cut locus, *cf.* my35, my36, used in [Be77] hold for every closed analytic surface.  $\square$

Given a critical point  $x$  of index 1 of  $F$ , there exists a unique (simple) geodesic loop  $\bar{\gamma}_x$  of index 0 based at  $x_0$  formed of two minimizing arcs,  $\bar{\gamma}'_x$  and  $\bar{\gamma}''_x$ , joining  $x_0$  to  $x$ . Conversely, if a (simple) geodesic loop of index 0 based at  $x_0$  is formed of two minimizing arcs joining  $x_0$  to  $x$ , the point  $x$  is a critical point of index 1. See the two figures in Section 3.

For technical reasons, it would be convenient if  $F$  had only one critical point on each critical level. To achieve this condition, we can raise some critical points of  $F$  as in [Ma02, Theorem 2.34]. More precisely, for every  $\varepsilon \in ]0, \text{inj}(M)/2[$  small enough, we can approximate  $F$  by a topological Morse function  $f$  on  $M$  preserving the critical points with only one critical point on each critical level and with  $x_0$  as its only local minimum, such that

- $f(x_0) = 0$ ;
- $\|f - F\| < \varepsilon$  on  $M$ ;
- $f$  is  $(1 + \varepsilon)$ -Lipschitz.

We can also assume that for every critical point  $x$  of index 1, with  $y := f(x)$ , the loop  $\bar{\gamma}_x$  satisfies the following:

- $\bar{\gamma}_x$  lies in  $f^{-1}([0, y])$ ;
- $\bar{\gamma}_x$  cannot be deformed into  $f^{-1}([0, y])$  in  $f^{-1}([0, y])$ ;

as well as

$$\text{length}(\bar{\gamma}_x \cap f^{-1}([t, \infty[)) \leq 2(y - t) + 2\varepsilon, \quad (4.1)$$

for every  $t \in [0, y]$ . The height of  $\bar{\gamma}_x$ , defined as  $\max f(\bar{\gamma}_x)$ , is equal to  $f(x) = y$ . Note that there are finitely many loops  $\bar{\gamma}_x$ .

**Definition 4.2.** Using the notations and definitions of Section 3, we say that the loop  $\bar{\gamma}_x$  is of type I or II if the critical point  $x$  is of type I or II. Furthermore, if  $x$  is of type II, the two trajectories  $\bar{\gamma}'_x$  and  $\bar{\gamma}''_x$  pass through the two cylinders of  $U_r(Z_x) \cap f^{-1}(]y - r, y])$  for  $r$  small enough. See the two figures in Section 3.

Let us construct by induction a “short” system  $\bar{\Gamma}$  of loops  $\bar{\gamma}_1, \dots, \bar{\gamma}_n$  based at  $x_0$ . The loop  $\bar{\gamma}_1$  is the noncontractible loop with the least height among the  $\bar{\gamma}_x$ 's. We define by induction  $\bar{\gamma}_i$  as the loop with the least height among the  $\bar{\gamma}_x$ 's whose homotopy class does not lie in the subgroup  $G_{i-1}$  generated by  $\bar{\gamma}_1, \dots, \bar{\gamma}_{i-1}$  in  $\pi_1(M, x_0)$ . Here, by definition,  $G_0$  is the trivial subgroup of  $\pi_1(M, x_0)$ . The loop  $\bar{\gamma}_i$  passes through two critical points, namely  $x_0$  and  $\bar{x}_i$ . Its height  $\max f(\bar{\gamma}_i)$  is equal to  $\bar{y}_i := f(\bar{x}_i)$ .

**Proposition 4.3.**

- i) *The system  $\bar{\Gamma}$  generates  $\pi_1(M, x_0)$ . In particular,  $\bar{\Gamma}$  contains at least  $2g$  loops;*
- ii) *The system  $\bar{\Gamma}$  contains at least  $g$  loops of type I;*
- iii) *If the homotopy class of a piecewise smooth loop  $\gamma$  based at  $x_0$  does not lie in  $G_{i-1}$ , then  $\max f(\gamma) \geq \max f(\bar{\gamma}_i)$ .*

*Proof.* Suppose that there exists a piecewise smooth loop  $\gamma$  based at  $x_0$  whose homotopy class does not lie in  $G_{i-1}$ . Using a height decreasing deformation of  $\gamma$  if necessary, we can assume that  $\gamma$  passes through a critical point  $x$  of index 1 such that

$$\max f(\gamma) = f(x) = y. \quad (4.2)$$

Thus, we can take  $\gamma$  (passing through a critical point  $x$  of index 1 satisfying (4.2)) so that its height is minimal.

The loop  $\bar{\gamma}_x$  passing through the critical point  $x$  of index 1 such that  $f(x)$  agrees with the height of  $\gamma$  satisfies the following.

**Lemma 4.4.** *The homotopy class of  $\bar{\gamma}_x$  does not lie in  $G_{i-1}$ .*

*Proof.* The loop  $\gamma$  passes finitely many times through  $x$ . Thus, we can take  $\gamma$  so that it passes a minimal number of times through  $x$ .

Define  $t_0 = \inf\{t \mid \gamma(t) = x\}$ . Either  $\gamma|_{[0,t_0]} \cup \bar{\gamma}'_x$  or  $\gamma|_{[0,t_0]} \cup \bar{\gamma}''_x$  can be deformed into a loop  $\alpha$  of  $f([0, y])$ .

Suppose that the former case occurs (similar arguments work in the latter case). The height of  $\alpha$  is less than  $\gamma$ 's. Since the height of  $\gamma$  is minimal, the homotopy class of  $\alpha$  lies in  $G_{i-1}$ . Therefore, we can assume that  $\gamma|_{[0,t_0]}$  agrees with  $\bar{\gamma}'_x$ . Since  $\gamma$  passes a minimal number of times through  $x$ , the loop  $\bar{\gamma}''_x \cup \gamma|_{[t_0,1]}$  can be deformed into a loop of height at most  $f(x)$ , which passes fewer times through  $x$  than  $\gamma$  (one should examine separately the type I and II cases; see the two figures in Section 3). Therefore, the homotopy class of the loop  $\bar{\gamma}''_x \cup \gamma|_{[t_0,1]}$  lies in  $G_{i-1}$ . Thus, we can assume that  $\gamma|_{[t_0,1]}$  agrees with  $\bar{\gamma}''_x$ . That is,  $\gamma$  agrees with  $\bar{\gamma}_x$ . Hence the desired result.  $\square$

Let us now prove i). Recall that  $n$  is the number of elements of  $\bar{\Gamma}$ . Suppose that there exists a piecewise smooth loop  $\gamma$  based at  $x_0$  whose homotopy class does not lie in the subgroup  $G_n$  generated by  $\bar{\Gamma}$ . By Lemma 4.4, there exists a loop  $\bar{\gamma}_x$  whose homotopy class does not lie in  $G_n$ . Therefore,  $\overline{\gamma_{n+1}}$  exists (and agrees with  $\bar{\gamma}_x$ ), which is absurd. Hence i).

To establish iii), we argue by contradiction again. By Lemma 4.4, there exists a loop  $\bar{\gamma}_x$  with  $y = f(x) < \max f(\bar{\gamma}_i)$  whose homotopy class does not lie in  $G_{i-1}$ . This yields a contradiction with the definition of  $\bar{\gamma}_i$ . Hence iii).

From Lemma 3.3, the number of loops of type II in  $\bar{\Gamma}$  is at most  $g$ . Since  $\bar{\Gamma}$  has exactly  $2g$  loops (each of them of type I or II), the system  $\bar{\Gamma}$  contains at least  $g$  loops of type I.  $\square$

Let  $\Gamma$  be the system of the  $g$  shortest loops  $\gamma_1, \dots, \gamma_g$  of type I in  $\bar{\Gamma}$ . Denote by  $y_i = \max f(\gamma_i)$  the height of  $\gamma_i$ . Permuting the indices if necessary, we can assume that  $y_1 < y_2 < \dots < y_g$ .

Let  $t_0 = 0$ ,  $t_1 = y_1$  and  $t_k = y_1 + (k-1)\tau$  for  $k \geq 2$ , where

$$\tau := \frac{1}{6} \frac{\bar{\sigma}'_0(g-1)}{g-1} \text{sys}_0(M). \quad (4.3)$$

and

$$\bar{\sigma}'_0(t) := \frac{1}{64} \frac{t}{\ln(2+t)} \quad (4.4)$$

for every real  $t \geq 0$ .

**Remark 4.5.** The loop  $\gamma_1$  agrees with  $\bar{\gamma}_1$ . Further, since  $x_0$  lies in a systolic loop of  $M$ , the height of  $\bar{\gamma}_1$  is bounded by  $\frac{1}{2} \text{sys}(M) + \varepsilon$ . That is,  $t_1 \leq \frac{1}{2} \text{sys}(M) + \varepsilon$ .

Denote by  $\gamma_1^k, \dots, \gamma_{g_k}^k$  the loops of  $\Gamma$  such that

$$t_k \leq \max f(\gamma_i^k) < t_{k+1}$$

and let  $y_i^k = \max f(\gamma_i^k)$  be the height of  $\gamma_i^k$ .

Let  $K$  be the minimal integer  $k$  such that  $y_g < t_{k+1}$ . We have

$$\sum_{k=1}^K g_k = g. \quad (4.5)$$

A real  $t$  is said to be *generic* if the preimage  $f^{-1}(t)$  is formed of a (possibly empty) finite union of disjoint circles. Since  $f$  is a topological Morse function (with finitely many critical points), almost every  $t$  is generic.

For a generic  $t$  with  $t_k < t < t_{k+1}$  and  $1 \leq k \leq K - 1$ , we define the graph

$$\Gamma_t^k := f^{-1}(t) \cup \bigcup_{i=1}^{g_{k+1}} (\gamma_i^{k+1} \cap f^{-1}([t, +\infty])). \quad (4.6)$$

The graph  $\Gamma_t^k$  is endowed with the length structure induced by the inclusion in  $M$ . The metric graph obtained from  $\Gamma_t^k$  by removing the connected components homeomorphic to circles is an admissible graph, *cf.* Definition 2.4, noted  $\widehat{\Gamma}_t^k$  (assuming that it is nonempty). We have

$$\widehat{\Gamma}_t^k \subset \Gamma_t^k.$$

Note that the graphs  $\Gamma_t^k$  and  $\widehat{\Gamma}_t^k$  differ from the graphs defined in [Ko87, §6] (these latter may contain arcs of loops of type I). In particular, the fundamental group of  $\widehat{\Gamma}_t^k$  is a subgroup of  $\pi_1(M)$ , *cf.* Section 5.

The vertices of  $\widehat{\Gamma}_t^k$  agree with the intersection points of  $f^{-1}(t)$  and  $\bigcup_{i=1}^{g_{k+1}} \gamma_i^{k+1}$ . Thus, the graph  $\widehat{\Gamma}_t^k$  is 3-regular (*i.e.*, the valence of each vertex equals 3) with  $2g_{k+1}$  vertices and  $3g_{k+1}$  edges. Recall that  $\Gamma_t^k$  lies in  $f^{-1}([t, +\infty])$ . Hence, the first Betti number of  $\widehat{\Gamma}_t^k$  is equal to  $g_{k+1}$  plus the number of connected components of  $\widehat{\Gamma}_t^k$ , *cf.* (2.1). In particular, if  $\widehat{\Gamma}_t^k$  is nonempty, its first Betti number is at least  $1 + g_{k+1}$ .

## 5. FUNDAMENTAL GROUPS OF GRAPHS AND SURFACES

In this section, where we use the previous constructions and notations, we show that the fundamental group of  $\widehat{\Gamma}_t^k$  lies in the fundamental group of  $M$ .

First, we show the following.

**Lemma 5.1.** *No simple loop of  $\widehat{\Gamma}_t^k$  is contractible in  $M$  for every generic  $t$  with  $t_k < t < t_{k+1}$  and  $1 \leq k \leq K - 1$ . In particular,*

$$\text{sys}(\widehat{\Gamma}_t^k) \geq \text{sys}(M).$$

*Proof.* We argue by contradiction. Let  $\gamma$  be a simple loop of  $\widehat{\Gamma}_t^k$  contractible in  $M$ . The loop  $\gamma$  bounds a disk  $D$  in  $M$ .

The function  $f$  admits no local minimum in the interior of  $D$ . Otherwise, such a local minimum agrees with the unique local minimum of  $f$ , which is  $x_0$ . In this case,  $D$  contains  $f^{-1}([0, t])$  and so  $\gamma_1$  since  $\max f(\gamma_1) = t_1 < t$ . Thus,  $\gamma_1$  is contractible in  $M$ . Hence a contradiction.

Now, two cases may occur:

CASE I: The loop  $\gamma$  is not contained in  $f^{-1}(t)$ . Then,  $\gamma$  passes through an arc  $\gamma_{i_0} \cap f^{-1}(]t, +\infty[)$  of some loop  $\gamma_{i_0}$  of  $\Gamma$ . Changing the index if necessary, we can assume that  $\gamma$  passes through no arc of  $\gamma_i$  with  $y_i \geq y_{i_0}$  (recall that  $f$  has only one critical point on each critical level). By construction, the simple arc  $\gamma$  is composed of arcs  $\gamma_i \cap f^{-1}(]t, +\infty[)$  where  $\gamma_i \in \Gamma$  and of subarcs of  $f^{-1}(t)$ . Since  $\gamma$  is contractible in  $M$ , the loop  $\gamma_{i_0}$  is homotopic to a loop based at  $x_0$  lying in  $f^{-1}([0, y_{i_0}[)$ . Thus, the homotopy class of  $\gamma_{i_0}$  lies in a subgroup generated by loops of height less than  $\gamma_{i_0}$ 's. This is impossible by definition of  $\gamma_{i_0}$ , cf. Proposition 4.3.iii).

CASE II: The loop  $\gamma$  is contained in  $f^{-1}(t)$ . Since  $f$  has no local minimum in the interior of  $D$ , the disk  $D$  lies in  $f^{-1}([t, +\infty[)$ . There exists a loop  $\gamma_i$  of  $\Gamma$ , with  $y_i > t$ , which intersects  $\gamma$  (recall that  $\gamma$  lies in  $\widehat{\Gamma}_t^k$  and that no connected component of  $\widehat{\Gamma}_t^k$  is homeomorphic to a circle). The arc  $\gamma_i \cap f^{-1}([t, +\infty[)$  lies in  $D$ . Therefore, the loop  $\gamma_i$  is homotopic to a loop based at  $x_0$  lying in  $f^{-1}([0, t])$ . As previously, we derive a contradiction.  $\square$

Let us introduce some notations. Fix a generic real  $t$  with  $t_k < t < t_{k+1}$  and  $1 \leq k \leq K - 1$  such that  $\widehat{\Gamma}_t^k$  is nonempty. Every sufficiently small open tubular neighborhood  $N$  of a connected component  $\Lambda$  of  $\widehat{\Gamma}_t^k$  deformation retracts onto  $\Lambda$  through  $\{r_s\}$  with  $s \in [0, 1]$  such that  $r_0$  is the identity map on  $N$  and  $r_1^{-1}(p) \cap \partial N$  has three or two elements depending whether the point  $p$  of  $\Lambda$  is a vertex or not (recall that  $\widehat{\Gamma}_t^k$  is 3-regular; see the figure below). We set  $r = r_1$ .



The connected components of the boundary  $\partial N$  of  $N$  are formed of finitely many (disjoint) simple loops. Furthermore, since  $M$  is orientable and the first Betti number of  $\Lambda$  is at least two, the set  $\partial N$  has at least three connected components. Thus,  $N$  is homeomorphic to a sphere with at least three holes. In particular,  $N$  is not homeomorphic to a cylinder.

The following result will be useful in the proof of Proposition 5.4.

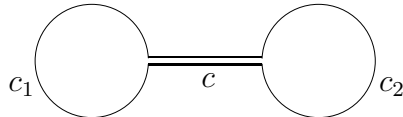
**Lemma 5.2.** *The simple loops forming the boundary  $\partial N$  of  $N$  are noncontractible in  $M$ .*

*Proof.* Suppose that there exists a simple loop  $\gamma$  of  $\partial N$  which is contractible in  $M$ . The image  $\bar{\gamma}$  of  $\gamma$  by the (deformation) retraction  $r$  is not simple, otherwise it would be noncontractible in  $M$  from Lemma 5.1 (and so would be  $\gamma$ ). Thus,  $r$  takes two points of  $\gamma$  to the same image. Since the graph  $\Lambda$  is 3-regular, the (deformation) retraction  $r$  takes two arcs of  $\gamma$  to the same edge  $c$  of  $\Lambda$ .

**Remark 5.3.** Changing the edge  $c$  if necessary, we will assume that the height of  $c$  is maximal among the edges of  $\Lambda$  onto which  $r$  sends two arcs of  $\gamma$ .

Removing  $c$  from  $\bar{\gamma}$  decomposes the loop  $\bar{\gamma}$  into two loops  $c_1$  and  $c_2$  with

$$\bar{\gamma} = c \cup c_1 \cup c^{-1} \cup c_2.$$



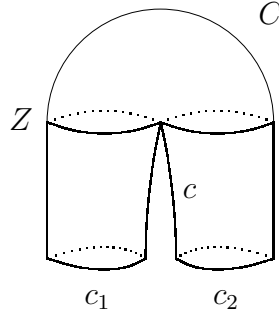
The loops  $c_1$  and  $c_2$  intersect the edge  $c$  only at its endpoints. Since  $\bar{\gamma}$  is the boundary of an open disk in  $M$ , the loops  $c_1$  and  $c_2$  are the boundaries of an open cylinder  $C$  in  $M$  containing  $c$ . In particular, the boundary  $\partial C$  of  $C$  agrees with  $c_1 \cup c_2$ .

The arcs  $c$  and  $c_i$  have different heights for  $i = 1, 2$ . Otherwise, their common height would be  $t$  (recall that  $f$  has at most one critical point

on each level set) and  $c \cup c_i$  would be contained in  $f^{-1}(t)$ , which is impossible since  $f^{-1}(t)$  is formed of a union of disjoint simple loops. The same argument also shows that  $c_1$  and  $c_2$  have different heights, unless they both lie in  $f^{-1}(t)$ .

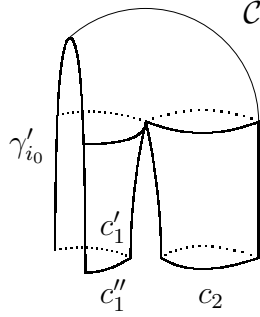
Now, we consider two cases.

CASE I: Assume that the height of  $c$  is greater than  $c_1$ 's and  $c_2$ 's. Note that  $y_c := \max f(c) > t$ . The connected component  $Z$  of  $f^{-1}(y_c)$  intersecting  $c$  lies in the cylinder  $C$  and is disjoint from  $\partial C$ . Thus, cutting  $C$  along the two noncontractible simple loops forming the component  $Z$  gives rise to two cylinders, at the bottom, and one disk, at the top (see figure below).



More precisely,  $f^{-1}([0, y_c]) \cap C$  is composed of two cylinders whose boundary components agree with the connected components of  $\partial C$  and the simple loops forming  $Z$ . We derive that the loop  $\gamma_c$  of  $\Gamma$  with  $\max f(\gamma_c) = y_c$ , which agrees with  $c$  in the neighborhood of  $f^{-1}(y_c)$ , *cf.* (4.6), is of type II. Hence a contradiction since the loops of  $\Gamma$  are of type I.

CASE II: Assume that the height of  $c_1$  is greater than  $c$ 's and  $c_2$ 's. In particular,  $\max f(c_1) > t$ . There exists a loop  $\gamma_{i_0}$  of  $\Gamma$ , with the same height as  $c_1$ , such that  $c_1 \cap \gamma_{i_0}$  agrees with the arc  $\gamma_{i_0} \cap f^{-1}([t, +\infty[)$ . The loop  $\gamma_{i_0}$  is formed of two arcs,  $\gamma'_{i_0}$  and  $\gamma''_{i_0}$ , of the same height, arising from  $x_0$  and ending at the same endpoint. From Remark 5.3, the loop  $\bar{\gamma}$  passes only once through  $c_1 \cap \gamma_{i_0}$ . Let  $c'_1$  and  $c''_1$  be the arcs of  $c_1 \setminus \gamma_{i_0}$  (possibly reduced to points) joining  $\gamma'_{i_0}$  to  $c$  and  $\gamma''_{i_0}$  to  $c$ .



The loop  $\gamma_{i_0}$  is homotopic to

$$(\gamma'_{i_0} \setminus c_1) \cup c'_1 \cup c'^{-1}_1 \cup (c_1 \cap \gamma_{i_0}) \cup c''_1 \cup c''^{-1}_1 \cup (\gamma''_{i_0} \setminus c_1),$$

which agrees with

$$(\gamma'_{i_0} \setminus c_1) \cup c'_1 \cup c_1 \cup c''^{-1}_1 \cup (\gamma''_{i_0} \setminus c_1).$$

By assumption, the loop  $c_1$  is homotopic to the loop  $c \cup c_2 \cup c^{-1}$ . Therefore,  $\gamma_{i_0}$  is homotopic to the loop

$$\alpha_{i_0} = (\gamma'_{i_0} \setminus c_1) \cup c'_1 \cup c \cup c_2 \cup c^{-1} \cup c''_1 \cup (\gamma''_{i_0} \setminus c_1).$$

We derive a contradiction from Proposition 4.3.iii) since the height of  $\alpha_{i_0}$  is less than  $\gamma_{i_0}$ 's.  $\square$

Let us recall some facts about the disk flow. Denote by  $M_0$  the surface  $M$  endowed with a fixed hyperbolic metric. The disk flow defined in [HS94] (see also [Sa04] for a similar flow) deforms a piecewise smooth loop  $\gamma$  of  $M_0$  through  $\gamma_t$  with  $t \geq 0$ . Let  $\mathcal{C}$  be a finite collection of piecewise smooth loops in  $M_0$ . Throughout the disk flow, the loops of  $\mathcal{C}$  satisfy the following:

- simple loops remain simple;
- disjoint loops remain disjoint;
- for every  $\gamma \in \mathcal{C}$ , the family  $\gamma_t$  either disappears (*i.e.*,  $\gamma_t$  converges to a point) in finite time or converges to a (unique) non-contractible geodesic loop of  $M_0$  as  $t \rightarrow \infty$ .

Using this flow and Lemma 5.2, we can prove the main result of this section.

**Proposition 5.4.** *The inclusion  $i : \Lambda \hookrightarrow M$  of every connected component of  $\widehat{\Gamma}_t^k$  induces a monomorphism between the fundamental groups. That is,*

$$i_* : \pi_1(\Lambda) \longrightarrow \pi_1(M)$$

*is injective.*

*Proof.* The boundary  $\partial N$  of  $N \subset M_0$  decomposes into simple loops  $c_1, \dots, c_k$ . We can assume that the normal vectors of the two-sided loops  $c_i$  point toward  $N$ . Consider a piecewise smooth loop  $\gamma$  of  $\Lambda \subset N$  contractible in  $M$  and apply the disk flow to  $c_1, \dots, c_k$  and  $\gamma$ .

The loop  $c_i$  converges to the unique noncontractible geodesic loop of  $M_0$  freely homotopic to  $c_i$  through the disk flow noted  $c_i^t$ . For every  $t \geq 0$ , the loops  $c_i^t$ , which are simple and disjoint as are the loops  $c_i$ , bound an open set  $N_t \simeq N$  toward which point the normal vectors of the two-sided loops  $c_i^t$ .

On the other hand, the loops  $\gamma_t$  remain disjoint from the loops  $c_i^t$  through the disk flow. In particular, the loops  $\gamma_t$  lie in  $N_t$ . Note that there exists a homeomorphism between  $N_t$  and  $N$  which takes  $c_i^t$  to  $c_i$  and the free homotopy class of  $\gamma_t$  to  $\gamma$ 's.

Since  $\gamma$  is contractible in  $M$ , the flow  $\gamma_t \subset N_t$  disappears in finite time. Therefore,  $\gamma_t$  is contractible in  $N_t$  for some  $t$ . This implies that  $\gamma$  is contractible in  $N$  and so in  $\Lambda$  since  $N$  deformation retracts onto  $\Lambda$ . Thus, the homomorphism  $i_* : \pi_1(\Lambda) \rightarrow \pi_1(M)$  is injective.  $\square$

We immediately deduce the following.

**Corollary 5.5.** *We have*

$$\text{sys}_0(\widehat{\Gamma}_t^k) \geq \text{sys}_0(M). \quad (5.1)$$

**Remark 5.6.** Contrary to  $\widehat{\Gamma}_t^k$ , the fundamental groups of the graphs defined in [Ko87] are not necessarily isomorphic subgroups of  $\pi_1(M)$ . Without this property, it is still possible to bound the systole of these graphs, cf. Lemma 5.1, but a bound similar to (5.1) does not hold in general.

## 6. PROOF OF THE MAIN THEOREM

Using the previous notations and results, we show the main result of this article following [Ko87].

*Proof of Main Theorem 1.3.* Since every smooth metric can be approximated by a real analytic one and since the area and the systole are continuous on the space of all metrics, we can assume that  $M$  is a real analytic Riemannian surface of genus  $g \geq 2$ . Keeping the same notations as in the previous sections, we can also assume that  $\text{sys}(M) < \frac{1}{4} \text{sys}_0(M)$ , otherwise the inequality (1.3) yields the result.

Let  $\gamma_1^1$  and  $\gamma_1^2$  be two loops of  $\Gamma$  based at  $x_0$  such that  $\max f(\gamma_1^1) = t_1 \leq \frac{1}{2} \text{sys}(M) + \varepsilon$  and  $\max f(\gamma_1^2) < t_1 + \tau$ , cf. Remark 4.5. The commutator of the homotopy classes of  $\gamma_1^1$  and  $\gamma_1^2$  can be represented by a loop  $c$  of length less than  $4t_1 + 2\tau$ . As  $t_1 < \frac{3}{4} \text{sys}(M) < \frac{3}{16} \text{sys}_0(M)$

and  $\tau < \frac{1}{8} \text{sys}_0(M)$ , the length of  $c$  is less than  $\text{sys}_0(M)$ . Thus,  $c$  is contractible and the homotopy classes of  $\gamma_1^1$  and  $\gamma_1^2$  commute. Since the centralizer of every nontrivial element of  $\pi_1(M)$  is isomorphic to  $\mathbb{Z}$  (recall that  $g \geq 2$ ), the homotopy classes of  $\gamma_1^1$  and  $\gamma_1^2$  are proportional. Furthermore, the homotopy class of a simple loop on the orientable surface  $M$  is indivisible. Therefore, the two noncontractible loops  $\gamma_1^1$  and  $\gamma_1^2$  of  $\Gamma$  agree. Thus, the index  $g_1$  defined in Section 4 is equal to 1, *i.e.*,

$$g_1 = 1.$$

Let  $t$  be generic with  $t_k < t < t_{k+1}$  and  $1 \leq k \leq K - 1$ . If  $\widehat{\Gamma}_t^k$  is nonempty, *i.e.*,  $g_{k+1} > 0$ , then  $\widehat{\Gamma}_t^k$  is an admissible graph, *cf.* Definition 2.4 and Section 4, with first Betti number at least  $1 + g_{k+1} \geq 2$ . Thus, since  $\sigma'_0$  is nondecreasing, *cf.* (2.4), we have

$$\text{length}(\widehat{\Gamma}_t^k) \geq \sigma'_0(1 + g_{k+1}) \text{sys}_0(\widehat{\Gamma}_t^k).$$

From Proposition 2.5, we derive that  $\sigma'_0(1 + g_{k+1}) \geq \bar{\sigma}'_0(g_{k+1})$ , *cf.* (4.4), where

$$\bar{\sigma}'_0(g_{k+1}) = \frac{1}{64} \frac{g_{k+1}}{\ln(2 + g_{k+1})}. \quad (6.1)$$

Hence, from the inclusion  $\widehat{\Gamma}_t^k \subset \Gamma_t^k$  and Corollary 5.5, we have

$$\text{length}(\Gamma_t^k) \geq \bar{\sigma}'_0(g_{k+1}) \text{sys}_0(M).$$

Furthermore, this inequality holds when  $g_{k+1}$  vanishes.

Thus, from the definition of  $\Gamma_t^k$ , *cf.* (4.6), we have

$$\text{length } f^{-1}(t) + \sum_{i=1}^{g_{k+1}} \text{length}(\gamma_i^{k+1} \cap f^{-1}([t, +\infty[)) \geq \bar{\sigma}'_0(g_{k+1}) \text{sys}_0(M)$$

Combined with the estimate (4.1), we derive

$$\text{length } f^{-1}(t) + \sum_{i=1}^{g_{k+1}} 2(t_i^{k+1} - t) + 2\varepsilon g_{k+1} \geq \bar{\sigma}'_0(g_{k+1}) \text{sys}_0(M)$$

where  $t_i^{k+1} = \max f(\gamma_i^{k+1})$ . Note that  $t_k < t < t_{k+1} \leq t_i^{k+1} < t_{k+2}$ .

Integrating from  $t_k$  to  $t_{k+1}$  for  $1 \leq k \leq K - 1$ , we obtain

$$\int_{t_k}^{t_{k+1}} \text{length } f^{-1}(t) dt + 2 \sum_{i=1}^{g_{k+1}} \int_{t_k}^{t_{k+1}} (t_i^{k+1} - t) dt + 2\varepsilon g_{k+1} \tau \geq \bar{\sigma}'_0(g_{k+1}) \text{sys}_0(M) \tau$$

Since  $\int_{t_k}^{t_{k+1}} t_i^{k+1} - t dt \leq \frac{3}{2} \tau^2$ , we have

$$\int_{t_k}^{t_{k+1}} \text{length } f^{-1}(t) dt \geq \bar{\sigma}'_0(g_{k+1}) \text{sys}_0(M) \tau - 3g_{k+1} \tau^2 - 2\varepsilon g_{k+1} \tau.$$

Now, apply the coarea formula to the  $(1 + \varepsilon)$ -Lipschitz function  $f$ , *cf.* [Fe69, 3.2.11]. One obtains

$$\text{Area}(M) \geq \frac{1}{1 + \varepsilon} \int_0^\infty \text{length } f^{-1}(t) dt.$$

Hence,

$$\begin{aligned} (1 + \varepsilon) \text{Area}(M) &\geq \sum_{k=1}^{K-1} \int_{t_k}^{t_{k+1}} \text{length } f^{-1}(t) dt \\ &\geq \left( \sum_{k=1}^{K-1} \bar{\sigma}'_0(g_{k+1}) \right) \text{sys}_0(M)\tau - 3 \left( \sum_{k=1}^{K-1} g_{k+1} \right) \tau^2 - 2\varepsilon \left( \sum_{k=1}^{K-1} g_{k+1} \right) \tau \\ &\geq \left( \sum_{k=2}^K \bar{\sigma}'_0(g_k) \right) \text{sys}_0(M)\tau - 3 \left( \sum_{k=2}^K g_k \right) \tau^2 - 2\varepsilon \left( \sum_{k=2}^K g_k \right) \tau \end{aligned}$$

Since  $g_1 = 1$ , the inequality (4.5) yields

$$\sum_{k=2}^K g_k = g - 1.$$

Furthermore, for every  $p, q \in \mathbb{N}$ ,

$$\begin{aligned} \bar{\sigma}'_0(p + q) &= \frac{1}{64} \left( \frac{p}{\ln(2 + p + q)} + \frac{q}{\ln(2 + p + q)} \right) \\ &\leq \frac{1}{64} \frac{p}{\ln(2 + p)} + \frac{1}{64} \frac{q}{\ln(2 + q)} \\ &\leq \bar{\sigma}'_0(p) + \bar{\sigma}'_0(q) \end{aligned}$$

In particular,

$$\sum_{k=2}^K \bar{\sigma}'_0(g_k) \geq \bar{\sigma}'_0 \left( \sum_{k=2}^K g_k \right) = \bar{\sigma}'_0(g - 1).$$

Hence,

$$(1 + \varepsilon) \text{Area}(M) \geq \bar{\sigma}'_0(g - 1) \text{sys}_0(M)\tau - 3(g - 1)\tau^2 - 2\varepsilon(g - 1)\tau.$$

Passing to the limit as  $\varepsilon \rightarrow 0$ , we obtain

$$\begin{aligned} \text{Area}(M) &\geq \bar{\sigma}'_0(g - 1) \text{sys}_0(M)\tau - 3(g - 1)\tau^2 \\ &\geq \frac{1}{12} \frac{\bar{\sigma}'_0(g - 1)^2}{g - 1} \text{sys}_0(M)^2 \end{aligned}$$

since  $\tau = \frac{1}{6} \frac{\bar{\sigma}'_0(g-1)}{g-1} \text{sys}_0(M)$ , cf. (4.3). Therefore,

$$\text{Area}(M) \geq \frac{1}{2^{16}} \frac{g-1}{(\ln(1+g))^2} \text{sys}_0(M)^2.$$

Thus, the inequality (1.7) holds for  $g \geq 2$  with  $C = 2^{-18}$ . Note that the value of the constant  $C$  can be improved, especially for large values of  $g$ .  $\square$

#### REFERENCES

- [BB05] Babenko, I.; Balacheff, F.: Géométrie systolique des sommes connexes et des revêtements cycliques, *Math. Ann.* **333** (2005) 157–180.
- [B04] Balacheff, F.: Sur des problèmes de la géométrie systolique, *Sémin. Théor. Spectr. Géom. Grenoble* **22** (2004) 71–82.
- [BCIK] Bangert, V.; Croke, C.; Ivanov, I.; Katz, M.: Boundary case of equality in optimal Loewner-type inequalities, *Trans. Amer. Math. Soc.* **359** (2007) 1–17.
- [Ba86] Bavard, C.: Inégalité isosystolique pour la bouteille de Klein, *Math. Ann.* **274** (1986) 439–441.
- [Ba93] Bavard, C.: L’aire systolique conforme des groupes cristallographiques du plan, *Ann. Inst. Fourier* **43** (1993) 815–842.
- [Be77] Berger, M.: Volume et rayon d’injectivité dans les variétés riemanniennes de dimension 3, *Osaka J. Math.* **14** (1977) 191–200.
- [Be93] Berger, M.: Systoles et applications selon Gromov, Séminaire Bourbaki, Exp. 711, *Astérisque* **216** (1993) 279–310.
- [BS02] Bollobàs, B.; Szemerédi, E.: Girth of sparse graphs, *J. Graph Theory* **39** (2002) 194–200.
- [BT97] Bollobàs, B.; Thomason, A.: On the girth of Hamiltonian weakly pancyclic graphs, *J. Graph Theory* **26** (1997) 165–173.
- [BS94] Buser, P.; Sarnak, P.: On the period matrix of a Riemann surface of large genus. With an appendix by J. H. Conway and N. J. A. Sloane. *Invent. Math.* **117** (1994), no. 1, 27–56.
- [CK03] Croke, C.; Katz, M.: Universal volume bounds in Riemannian manifolds, *Surveys in Differential Geometry VIII*, Lectures on Geometry and Topology held in honor of Calabi, Lawson, Siu, and Uhlenbeck at Harvard University, May 3–5, 2002, edited by S.T. Yau (Somerville, MA: International Press, 2003.) pp. 109–137. See [arXiv:math.DG/0302248](https://arxiv.org/abs/math/0302248)
- [Fe69] Federer, H.: Geometric measure theory. *Grundlehren der mathematischen Wissenschaften*, **153**. Springer–Verlag, Berlin, 1969.
- [Fi40] Fiala, F.: Le problème des isopérimètres sur les surfaces ouvertes à courbure positive, *Comment. Math. Helv.* **13** (1941) 293–346.
- [Gr83] Gromov, M.: Filling Riemannian manifolds, *J. Diff. Geom.* **18** (1983) 1–147.
- [Gr96] Gromov, M.: Systoles and intersystolic inequalities. *Actes de la Table Ronde de Géométrie Différentielle (Luminy, 1992)*, 291–362, Sémin. Congr., vol. 1, Soc. Math. France, Paris, 1996.  
[www.emis.de/journals/SC/1996/1/ps/smf\\_sem-cong\\_1\\_291-362.ps.gz](http://www.emis.de/journals/SC/1996/1/ps/smf_sem-cong_1_291-362.ps.gz)

- [Gr99] Gromov, M.: Metric structures for Riemannian and non-Riemannian spaces. *Progr. in Mathematics* **152**, Birkhäuser, Boston, 1999.
- [HS94] Hass, J.; Scott, P.: Shortening curves on surfaces, *Topology* **33** (1994) 25–43.
- [He82] Hebda, J.: Some lower bounds for the area of surfaces, *Invent. Math.* **65** (1982) 485–490.
- [IK04] Ivanov, S.; Katz, M.: Generalized degree and optimal Loewner-type inequalities, *Israel J. Math.* **141** (2004) 221–233.
- [KSV] Katz, M.; Schaps, M.; Vishne, U.: Logarithmic growth of systole of arithmetic Riemann surfaces along congruence subgroups, preprint available at arXiv:math.DG/0505007
- [KS05] Katz, M.; Sabourau, S.: Entropy of systolically extremal surfaces and asymptotic bounds, *Ergod. Th. Dynam. Sys.*, **25** (2005), 1209–1220.
- [KS06a] Katz, M.; Sabourau, S.: Hyperelliptic surfaces are Loewner, *Proc. Amer. Math. Soc.* **134** (2006), no. 4, 1189–1195.
- [KS06b] Katz, M.; Sabourau, S.: An optimal systolic inequality for CAT(0) metrics in genus two, *Pacific J. Math.* **227** (2006), no. 1, 155–176.
- [Ko87] Kodani, S.: On two-dimensional isosystolic inequalities, *Kodai Math. J.* **10** (1987), no. 3, 314–327.
- [Ma02] Matsumoto, Y.: An introduction to Morse theory, Translations of Mathematical Monographs, **208**, Amer. Math. Soc., Providence, RI, 2002
- [Mo59] Morse, M.: Topologically non-degenerate functions on a compact  $n$ -manifold  $M$ , *J. Analyse Math.* **7** (1959) 189–208.
- [Mo75] Morse, M.: Topologically nondegenerate functions, *Fund. Math.* **88** (1975) 17–52.
- [My35] Myers, S. B.: Connections between differential geometry and topology, I. Simply connected surfaces, *Duke Math.* **1** (1935), no. 3, 376–391.
- [My36] Myers, S. B.: Connections between differential geometry and topology, II. Closed surfaces, *Duke Math.* **2** (1936), no. 1, 96–102.
- [Pu52] Pu, P. M.: Some inequalities in certain nonorientable Riemannian manifolds, *Pacific J. Math.* **2** (1952) 55–71.
- [Sa04] Sabourau, S.: Filling radius and short closed geodesics of the 2-sphere, *Bull. Soc. Math. France* **132** (2004) 105–136.
- [Sa06a] Sabourau, S.: Entropy and systoles on surfaces, *Ergod. Th. Dynam. Sys.*, **26** (2006), no. 5, 1653–1669.
- [Sa06b] Sabourau, S.: Systolic volume and minimal entropy of aspherical manifolds, *J. Diff. Geom.*, **74** (2006), no. 1, 155–176.

LABORATOIRE DE MATHÉMATIQUES ET PHYSIQUE THÉORIQUE, UNIVERSITÉ DE TOURS, PARC DE GRANDMONT, 37400 TOURS, FRANCE

*E-mail address:* sabourau@lmpt.univ-tours.fr