# New Length Bounds for Cycle Bases ⋆

Michael Elkin [a] Christian Liebchen [b] Romeo Rizzi [c,∗]

[a] *Department of Computer Science, Ben-Gurion University of the Negev, Beer-Sheva, Israel*

[b] *Technische Universität Berlin, Institut für Mathematik, Combinatorial Optimization and Graph Algorithms, Straße des 17. Juni 136, D-10623 Berlin, Germany*

[c] *DIMI, Università di Udine, Italy*

**Abstract**

Based on a recent work by Abraham, Bartal and Neiman (2007), we construct a strictly fundamental cycle basis of length $O(n^2)$ for any unweighted graph, whence proving the conjecture of Deo et al. (1982).

For weighted graphs, we construct cycle bases of length $O(W \cdot \log n \log\log n)$, where $W$ denotes the sum of the weights of the edges. This improves the upper bound that follows from the result of Elkin et al. (2005) by a logarithmic factor and, for comparison from below, some natural classes of large girth graphs are known to exhibit minimum cycle bases of length $\Omega(W \cdot \log n)$.

We achieve this bound for weighted graphs by not restricting ourselves to strictly fundamental cycle bases—as it is inherent to the approach of Elkin et al.—but rather also considering weakly fundamental cycle bases in our construction. This way we profit from some nice properties of Hierarchically Well-Separated Trees that were introduced by Bartal (1998).

## 1 Introduction

We consider a simple undirected 2-connected graph $G = (V, E)$ with $n$ nodes and $m$ edges. A non-negative weight function $w$ may be defined over the edge set $E$. Otherwise we are in the unweighted case, and $w \equiv 1$ is implicitly assumed. In

⋆ Supported by the DFG Research Center MATHEON in Berlin and by the Israeli Academy of Science, grant 483/06.

∗ Corresponding Author.
*Email address:* romeo.rizzi@dimi.uniud.it (Romeo Rizzi).

the weighted case we denote by $W$ the sum of the weights of all the edges. In the unweighted case $W = m$. A *circuit* $C$ of $G$ is a connected subgraph of $G$, in which each vertex has either degree two or zero. Sometimes, we may refer to $C$ only through its edge set. The *incidence vector* $\gamma_C \in \{0,1\}^E$ of a circuit $C$ is the characteristic vector of its edge set. The weight $w(C)$ of a circuit $C$ is defined as the sum of the weights of its edges.

**Cycle Bases.** The *cycle space* $\mathcal{C}(G)$ of $G$ is the linear vector subspace of GF$(2)^E$ which is spanned by the incidence vectors of the circuits of $G$. It is well known that $\nu := \dim(\mathcal{C}(G)) = m - n + 1$.

A *cycle basis* $\mathcal{B}$ of $G$ is a set of $\nu$ circuits of $G$ which span $\mathcal{C}(G)$. The weight $w(\mathcal{B})$ of a cycle basis $\mathcal{B}$ is defined as $w(\mathcal{B}) = \sum_{C \in \mathcal{B}} w(C)$. In the *Minimum Cycle Basis (MCB) Problem*, we seek a cycle basis of minimum weight. These turn useful in many applications, see [Hor87] and references therein.

Given a spanning tree $T$ of $G$, and a non-tree edge $e \in E \setminus T$, the unique circuit in $T \cup \{e\}$ is denoted by $C_T(e)$ and called the *fundamental circuit* induced by $e$ with respect to $T$. It is well known that the set of all the fundamental circuits w.r.t. a spanning tree $T$ is a cycle basis; these cycle bases have special and often desirable properties, and are called *strictly fundamental*. A cycle basis is called *weakly fundamental* if its circuits can be labeled as $C_1, C_2, \ldots, C_\nu$ such that

$$C_i \setminus (C_1 \cup \cdots \cup C_{i-1}) \neq \emptyset, \quad \text{for all } i = 2, \ldots, \nu. \tag{1}$$

Let $e_i \in C_i \setminus (C_1 \cup \cdots \cup C_{i-1})$. Notice that $\{e_1, \ldots, e_\nu\}$ are the co-tree arcs with respect to some spanning tree $T$ of $G$. In particular, the circuit $C_\nu$ is a fundamental circuit with respect to that spanning tree $T$.

**Metrics.** The *metric* $d_G$ over $V$ that is associated with a weighted connected graph $(G, w)$ is defined by the shortest-path distances in $G$ with respect to $w$. The *diameter* diam$(G)$ of $(G, w)$ is defined as

$$\text{diam}(G) := \max\{d_G(u, v) : \ u, v \in V\}. \tag{2}$$

Let $V' \subset V$ and consider the induced subgraph $G[V']$ of $G$. The *strong internal diameter of* $V'$ is diam$(G[V'])$. In contrast, the *weak diameter of* $V'$ is the maximum distance *in the original graph* $G$ between two vertices in $V'$. A metric $d$ on a set of elements $V$ is said to *dominate* another metric $d'$ on $V$ if for all $u, v \in V$ we have $d'(u, v) \leq d(u, v)$. In particular, given a family $V_1, \ldots, V_k$ of connected subsets of $V$, a metric $d$ dominates the strong internal diameters of $V_1, \ldots, V_k$ over $(G, w)$, if for every subset $V_i$, $i = 1, \ldots, k$, and pair of vertices $u, v \in V_i$,

$$d(u, v) \geq \text{diam}(G[V_i]) \ . \tag{3}$$

**Related Work.** Deo et al. ([DKP82]) conjecture that every unweighted graph has a strictly fundamental cycle basis of length $O(n^2)$. Since strictly fundamental cycle bases specialize general cycle bases, their conjecture may also serve as the first estimate on the length of an MCB. For unweighted graphs, Horton ([Hor87]) introduced a heuristic for computing a cycle basis of length $O(n^2)$. The running time of Horton's algorithm is $O(m^3 n)$. Later, it was shown in [Lie03] that the resulting cycle basis is always weakly fundamental. As this basis is not strictly fundamental in general, these results provide only a partial answer to Deo et al.'s conjecture.

Alon et al. ([AKPW95]) gave a more direct answer to Deo et al.'s conjecture. They prove that every weighted graph has a strictly fundamental cycle basis of length $O(W \cdot \exp(O(\sqrt{\log n \log\log n})))$, hereby proving Deo's conjecture in the case of graphs that are not too dense. Recently, this was improved by Elkin et al. ([EEST05]) to $O(W \cdot \log^2 n \log\log n)$. It is known that there exist graphs with large girth such that the length of their MCB is $\Omega(W \cdot \log n)$, see [Bol78]. Hence, for general graphs the bound of Elkin et al. is asymptotically tight up to a factor of $\Theta(\log n \log\log n)$.

Bartal ([Bar98]) showed that any metric can be approximated by a tree metric, which is defined on an auxiliary graph, called Hierarchically Well-Separated Tree (HST). More precisely, the resulting tree metric dominates the initial metric and overestimates it—in the deterministic case on average—by a factor of at most $\alpha(n) \in O(\log n \log\log n)$. Later, by introducing further techniques, this bound was improved to the tight $O(\log n)$ ([FRT03]).

In parallel, MCB algorithms that improve the running time of Horton's algorithm [Hor87] ($O(m^3 n)$) were developed. The presently fastest known MCB algorithm has running time $O(m^2 n / \log n + n^2 m)$ [MM07]. Also constant-factor approximation algorithms had been proposed [KMMP04,KMM07,MM07], e.g. a 2-approximation can be constructed in time $\tilde{O}(m^{3/2} n^{3/2}) + O(m^\omega)$, where $\omega < 2.376$ is the matrix multiplication exponent. However, these algorithms have in common that they do not provide any bounds on the absolute length of the MCB.

**Contribution.** We develop a way to profit from Bartal's techniques in the context of the Minimum Cycle Basis Problem. More specifically, using HSTs [Bar98] we construct a weakly fundamental cycle basis of length $O(W \cdot \log n \log\log n)$ for any weighted 2-connected undirected graph. Hereby, we improve the previously known best upper bound on the length of general minimum cycle bases by a factor of $\log n$. Also, due to the lower bound of $\Omega(W \cdot \log n)$, we conclude that our construction is tight up to a factor $\Theta(\log\log n)$. The running time of the algorithm that we are proposing here is bounded by the polynomial amount of time that Bartal's construction of a HST requires.

However, in [**?**] a much more efficient algorithm is given. It achieves the very same length bound and its running time is $O(n^2 \cdot \log^{c+1} n + \sum_{C \in \mathcal{B}} |C|)$, where $\{C \mid C \in \mathcal{B}\}$ is the cycle basis constructed by our algorithm, and $c = O(1)$ is the universal constant that is implicit in the result of Cohen and Zwick [CZ01]. In particular, in the important case of sparse graphs, among the class of algorithms with computation time $o(n^3)$, hereby the best known approximation guarantee is obtained.

Furthermore, based on a recent work by Abraham, Bartal and Neiman [ABN07], we construct a strictly fundamental cycle basis of length $O(n^2)$ for any unweighted graph, whence proving the conjecture of Deo et al. [DKP82].

**Outline.** This paper is organized as follows. In Section 2 we overview the key properties of HSTs. Based on HSTs, in Section 3 we present the algorithm for computing a weakly fundamental cycle basis of length $O(W \cdot \log n \log\log n)$. The proof of the conjecture of Deo et al. [DKP82] is given in Section 4.

## 2 Techniques and Main Idea

In following the presentation of our algorithm, the reader should keep in mind the following simple—though useful—property of weakly fundamental cycle bases.

**Remark 1** *One way to construct a weakly fundamental cycle basis $\mathcal{B}$ is to define some spanning tree $F$ and provide an order of the co-tree edges $E \setminus F$, say $e_1, \ldots, e_\nu$, where $F = e_{\nu+1}, \ldots, e_m$. Then, the basic circuit $C$ that we associate with a co-tree edge $e_i \in E \setminus F$ has to contain $e_i$ and may use some edges of the tree $F$ and some of the co-tree edges $e_1, \ldots, e_{i-1}$. But $C$ must not contain any of the edges $e_{i+1}, \ldots, e_\nu$.*

The key to our algorithm for constructing a short weakly fundamental cycle basis will thus be the order in which we process the edges of $G$. And this order will be dictated by a decomposition of $G$, the HST. This object was introduced by Bartal ([Bar96]).

Consider a nested—or laminar—family $\mathcal{S}$ of connected subgraphs of $G$. We only consider such families in which the maximal element is the entire vertex set $V$, and all the singletons in $V$ appear as minimal elements in $\mathcal{S}$. For $U \in \mathcal{S}$ we denote by $G[U]$ its induced subgraph, and call it *cluster*.

We resort to the standard way of representing such a laminar family $\mathcal{S}$ by a rooted tree $T$: the nodes of $T$ are the sets in $\mathcal{S}$, and there is an arc $(U, U')$ between two sets $U$ and $U'$ in $\mathcal{S}$ iff $U' \subseteq U$ and $U' \subseteq U'' \subseteq U$ holds for no $U''$ in $\mathcal{S}$. Notice that $V$ is the root of $T$ and the leaves of $T$ are in 1,1-correspondence with the nodes in $V$. More generally, each node of $T$ is the disjoint union of its children. Such a tree is called a *$V$-tree*. In a $V$-tree $T$, the *least common ancestor* $\Lambda(u, v)$ of a pair of vertices $u \neq v$ is the vertex in the unique $uv$-path in $T$ that is closest to the root of $T$. Its associated cluster in $G$ is just $G[\Lambda(u, v)]$.

**Definition 2** *A* hierarchically well-separated tree (HST) *of a weighted connected graph $(G, w)$ is a $V$-tree $T$ with weights $c$ on the arcs, with the following properties:*

*(1) $c(U, U_1) = c(U, U_2)$ for any two arcs of $T$ with a common tail;*
*(2) $c(U', U'') \leq \frac{1}{2}c(U, U')$ for any two subsequent arcs $(U, U')$ and $(U', U'')$ in $T$;*
*(3) for each node $U$ of $T$, the induced graph $G[U]$ is connected.*

An HST $(T, c)$ induces a tree metric $d_{HST}$ over $V$, where $d_{HST}(u, v)$ is the distance between $u$ and $v$ in $(T, c)$, i.e. $d_{HST}(u, v) := c(P_{u,v})$, with $P_{u,v}$ being the unique $u, v$-path in $T$. We only consider HSTs in which $d_{HST}$ dominates $d_{G,w}$. Whenever this is the case we say that the HST dominates the metric of $(G, w)$ with respect to *weak diameters*. A stronger condition is to require that for every cluster $U$ of the HST, the metric $d_{G[U],w}$ of the induced graph $(G[U], w)$ of the cluster $U$ is dominated by the metric $d_{HST}$. Whenever this is the case we say that the HST dominates the metric of $(G, w)$ with respect to *strong diameters*.

We are interested in HSTs of low average stretch, that is, we need a bound on $\sum_{\{u,v\}\in E} d_{\mathrm{HST}}(u, v)$. Also, we require the strong internal diameters of the clusters to be dominated. Notice that in [Bar98, Thm. 10] the result that we are about to use has only been stated explicitly for weak diameters. However, since the HST in [Bar98] is built on top of a Hierarchical Partition Metric (HPM), the result is even true when considering the strong internal diameters.

**Theorem 3 (stated in [Bar98] for weak diameters)** *There exists a function $\alpha(n) \in \Theta(\log n \log\log n)$ such that for every weighted graph $(G, w)$ there exists a polynomial-time algorithm to construct an HST $T$ with metric $d_{HST}$ that dominates the metric of $(G, w)$ with respect to strong diameters that satisfies*

$$\sum_{\{u,v\}\in E} d_{HST}(u, v) \leq \alpha(n) \cdot W. \tag{4}$$

Recall that an alternative proof for this theorem is also provided in [**?**].

In the algorithm for computing a short weakly fundamental cycle basis that we are about to present in the next section, we will process the clusters of $T$ bottom-up, i.e. we will only start working in a cluster when we finished working in all of its descendents.

## 3 Computing a Short Weakly Fundamental Cycle Basis

Now we present our Algorithm 1 that for a given weighted graph $(G, w)$ of total weight $W$ constructs a weakly fundamental cycle basis of objective value at most

$$(16 \cdot \alpha(n) + 1) \cdot W, \quad \alpha \in O(\log n \log\log n), \tag{5}$$

where $\alpha$ refers to Bartal's function in Theorem 3.

The first line of Algorithm 1 is a call to Procedure MAKE-HST, providing a hierarchically Well-Separated Tree $T$ that dominates the strong internal diameters of its corresponding clusters of$(G, w)$, and satisfies the length-bound in (4) (Theorem 3). In principle, it was shown in [Bar98] how this could be performed in polynomial time. We refer to [**?**] for the details of this construction, as well as to [**?**] for a much more efficient construction.

Algorithm 1 resorts on a further external procedure: we assume that if $(G, w)$ is a weighted graph and $u$ is a node of $G$, then DIJKSTRA$((G, w), u)$ computes the $u$-rooted shortest path tree $S = e_1, e_2, \ldots, e_{n-1}$ in $(G, w)$; it is assumed that the edges of $S$ are given in the same order as they would be put into $S$ by the classical Dijkstra's algorithm ([Dij59]). In other words, the sequence $e_1, e_2, \ldots, e_s$ satisfies the following property. Consider a pair of edges $e = (v, w), e' = (v', w')$ in the sequence with $v$ (respectively, $v'$) being closer to the root $u$ of $S$ than $w$ (resp., $w'$). Suppose that $e$ belongs to the path connecting the root $u$ of the tree $S$ with the endpoint $v'$ of $e' = (v', w')$. For every pair of edges $e, e'$ as above it is required that the edge $e$ appears in the sequence before $e'$.

After the hierarchically well-separated tree $T$ and the corresponding metric $d_{\text{HST}}$ have been obtained, the actual iterative construction of the circuits to be put in the basis $\mathcal{B}$ can start. As anticipated in Remark 1, we will accompany the construction of $\mathcal{B}$ by also computing a spanning tree $F$ of $G$ plus an ordering of the non-tree edges $E(G) \setminus F$ of $G$. Thus, in our iterative process we start with $\mathcal{B} := F := Z := \emptyset$, and collect in the set $Z$ all the edges that were already processed according to the ordering that we are about to define. Observe that $Z$ will contain tree edges *and* non-tree edges. When we process an edge $e$ (and thus add $e$ to $Z$), this has one of the two possible results: Either $F \cup \{e\}$ is acyclic, then we add $e$ to $F$. Or $F \cup \{e\}$ contains some cycle, then we add to $\mathcal{B}$ a circuit through $e$ which only uses edges in $Z$, and which is sufficiently short. For simplicity, you may think of a shortest circuit through $e$ that only uses edges in $Z$.

We now give a description of the order in which we are going to process the edges of $G$. Processing the edges will be grouped according to the clusters of $T$. We will color a node $U \in V(T)$ *green*, if all the edges of its corresponding cluster $G[U]$ have been processed, i.e. $E(G[U]) \subseteq Z$. Otherwise, $U$ will have to remain colored *red*. At the start of the algorithm we may thus color in green all the leaves of $T$, and the algorithm terminates when the root of $T$ finally becomes green. The order by which we pass through the clusters of the HST will be bottom-up, i.e. we may only work in a node $U \in V(T)$, if all its descendents are colored green. Hence, when we start working in $G[U]$, some edges of $G[U]$ may have already been processed while working in clusters that correspond to descendents of $U$ in $T$. The other edges, i.e. $E(G[U]) \setminus Z$, are called *$U$-proper*. We process all the $U$-proper edges in a specific order which we specify in the very next paragraph. After doing so, we may color $U$ green.

The order in which we process the $U$-proper edges of a cluster $G[U]$ is as follows. We first compute Dijkstra's shortest path tree $S \subseteq E(G[U])$ in this cluster, rooted at some arbitrary vertex $u$ of $U$. During this procedure, whenever a $U$-proper edge $e$ gets added to $S$, we process this edge. Second, after all the edges of $S$ have been processed, i.e. $S \subseteq Z$, we process—in arbitrary order—all the $U$-proper edges in $E(G[U]) \setminus S$.

In practice, we suggest to compute the basic circuits which we add in Step 24 of Algorithm 1 as the shortest circuits through $e$ in $Z \cup \{e\}$. In general, this would lead to shorter bases. However, our asymptotic analysis would not improve.

For a cluster $G[U]$, let

$$\Delta(U) := \text{diam}_{\text{HST}}(G[U]) := \max\{d_{\text{HST}}(u, v),\ u, v \in U\}.$$

By Theorem 3, $d_{\text{HST}}$ dominates the metric with respect to strong diameters, in particular

$$\Delta(U) \geq \text{diam}(G[U]). \tag{6}$$

Moreover, for any two vertices $v_1, v_2 \in V(G)$ for which $\Lambda(v_1, v_2) = U$, we may bound $\Delta(U)$ as follows,

$$\Delta(U) \leq 2 \cdot \left(\frac{1}{2} d_{\text{HST}}(v_1, v_2) \sum_{k=0}^{\infty} \frac{1}{2^k}\right) \leq 2 \cdot d_{\text{HST}}(v_1, v_2), \tag{7}$$

where inside the brackets we are considering one path in the HST from $U$ to one of its leaves, and use Property (2) of Definition 2.

**Lemma 4** *Let $U$ be the cluster that is currently processed by Algorithm 1. Each circuit $C$ that we add to the basis $\mathcal{B}$ in Step 18 as the shortest circuit through $e_i = \{v_1, v_2\} \in S$ in $Z$ has weight $w(C)$ at most*

$$w(C) \leq 5 \cdot d_{HST}(v_1, v_2). \tag{8}$$

*Proof.* Here, $e_i \notin Z$, and thus $\Lambda(v_1, v_2) = U$. Hence we will first identify a bound on $w(C)$ in terms of $\Delta(U)$, and finally apply (7). The bound on $w(C)$ is obtained by identifying a (possibly different) cycle $C'$ in $Z$ and with $e_i \in C'$, and establishing

**Algorithm 1** `wfcb`

**Require:** A weighted input graph $(G, w)$.
**Ensure:** A weakly fundamental cycle basis $\mathcal{B}$ of $G$ with $w(\mathcal{B}) = O(W \cdot \log n \log \log n)$.

$(T, d_{\text{HST}}) := \text{MAKE-HST}(G, w)$; // see appendix
$\mathcal{B} := \emptyset$;
$F := \emptyset$; // becomes a spanning tree of $G$ related to $\mathcal{B}$ as in Remark 1
$Z := \emptyset$; // the set of edges already considered. In particular, $F \subseteq Z$ and each edge in $Z \setminus F$ belongs to some circuit in $\mathcal{B}$.
Color with *green* all the leaves of $T$, and with *red* all other nodes;
**while** some node of $T$ is *red* **do**
  Let $U$ be any *red* node of $T$ whose children are all *green*;
  Color $U$ with *green*;
  Let $u$ be an arbitrary vertex of $G[U]$;
  $S = \{e_1, \ldots, e_s\} := \text{DIJKSTRA}((G[U], w), u)$; // Recall that $G[U]$ is connected.
  **for** $i = 1$ to $s$ **do**
    **if** $e_i \in Z$ **then**
      // void — $e_i$ already processed in a subcluster of $U$.
    **else if** $F \cup \{e_i\}$ is acyclic **then**
      $F := F \cup \{e_i\}, \quad Z := Z \cup \{e_i\}$;
    **else**
      // $e_i \notin Z$, and $F \cup \{e_i\}$ contains some cycle
      Add to $\mathcal{B}$ the shortest circuit through the edge $e_i$ that only uses edges in $Z$;
      $Z := Z \cup \{e_i\}$;
    **end if**
  **end for**
  // By now we have $S \subseteq Z$.
  **for all** $e \in E(G[U]) \setminus Z$ **do**
    Add to $\mathcal{B}$ the unique circuit in $S \cup \{e\}$;
    $Z := Z \cup \{e\}$;
  **end for**
  // Here we have $E(G[U]) \subseteq Z$.
**end while**

$w(C') \leq 5 \cdot \Delta(U)$. The cycle $C'$ will be constructed out of two subpaths of $S$ plus a shortest path within a subcluster of $U$.

We assume without loss of generality that $v_1$ is closer to the root $u$ of the shortest-path tree $S$ in $G[U]$ than $v_2$. Denote by $U_2$ the child of $U$ in $T$ that contains $v_2$.

We are in the situation where $F \cup \{e_i\}$ contains some circuit $Q$ with $e_i \in Q$. Consider the cut $X \subset E(G[U])$ that corresponds to $(U_2, U \setminus U_2)$. Observe that $e_i \in X$. As every cycle has even intersection with any cut, and since $\emptyset \neq \{e_i\} \in Q \cap X$, we know that $|Q \cap X| \geq 2$, and in particular $|F \cap X| \geq 1$.[1]

But since edges in $X$ have precisely one endpoint in $U_2$, they cannot be contained in any subcluster of $U$ and thus are processed during $U$'s iteration of the **while**-loop. Hence, each element of $F \cap X$ is part of the shortest-path tree $S$ in $U$.

We compose the not necessarily simple cycle $C'$ out of the following three parts: First, the unique path $P_{ue_i} \subset S$ from $u$ to $e_i$. Second, one path $P_{uf} \subset S$ from $u$ to some edge $f$ such that $P_{uf} \cap X = \{f\} \neq \{e_i\}$, see Figure 1 for the location of $e$, $f$, $u$, and $U_2$. As both these paths are subpaths of the shortest-path tree $S$ in $G[U]$,

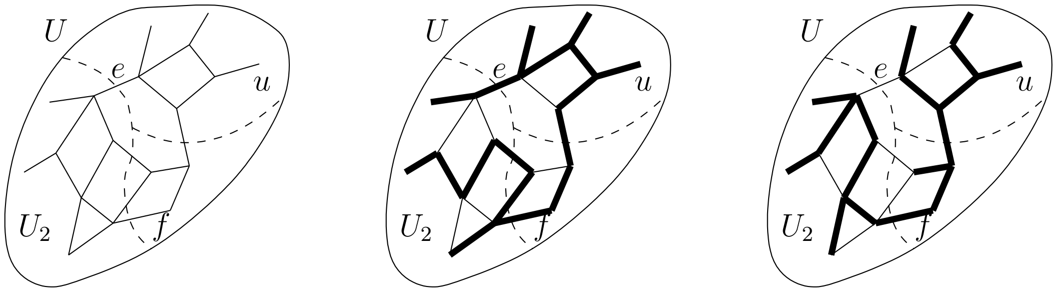


Fig. 1. Comparing for some cluster $U$ (on the left) the $u$-rooted shortest-path tree $S$ (in the middle) to the growing spanning tree $F$ of the input graph (on the right).

for $P \in \{P_{ue_i}, P_{uf}\}$ we know that $w(P) \leq \text{diam}(G[U]) \overset{(6)}{\leq} \Delta(U)$...

The third and last part of the cycle $C'$ is selected as a shortest path $P_{e_if}$ in $G[U_2]$ between the corresponding endpoints of $e_i$ and $f$. Such a path exists because any cluster is connected, and $P_{e_if}$ only uses edges in $Z$ because it stays within the cluster $U_2$ all of whose edges were already processed, cf. the comment in Line 27 of Algorithm 1.

To bound the length of $P_{e_if}$, recall the second property of an HST and consider the subgraph $G[U_2]$ of $G$: Our bound $\Delta(U_2)$ on the strong internal diameter of $G[U_2]$ is by (at least) a factor of $k$ smaller than the corresponding bound $\Delta(U)$ on the strong internal diameter of $G[U]$. Hence, we finally obtain $w(P_{e_if}) \leq \text{diam}(G[U_2]) \leq \Delta(U_2) \leq \frac{\Delta(U)}{2}$. Overall we obtain

$$
\begin{aligned}
w(C) &\leq w(C') \leq w(P_{ue_i}) + w(P_{uf}) + w(P_{e_if}) \\
&\leq 2 \cdot \Delta(U) + \frac{\Delta(U)}{2} = \frac{5}{2} \cdot \Delta(U) \overset{(7)}{\leq} 5 \cdot d_{\text{HST}}(v_1, v_2).
\end{aligned}
$$

□

**Lemma 5** *Let $U$ be the cluster that is currently processed by Algorithm 1. Each circuit $C$ that we add to the basis $\mathcal{B}$ in Step 24 for some non-tree edge $e = \{v_1, v_2\} \in$*

[1] One could even show by induction that in fact $|F \cap X| = 1$.

$E(G[U]) \setminus S$ *has weight* $w(C)$ *at most*

$$w(C) \leq 4 \cdot d_{HST}(v_1, v_2) + w(e). \tag{9}$$

*Proof.* We identify a cycle $C'$ that respects the length-bound in (9), but which is already a superset of the circuit $C$ that we add to the basis in Step 24. Again, from $e = \{v_1, v_2\} \notin Z$ we conclude that $U = \Lambda(v_1, v_2)$. Let $P_1$ be the unique $uv_1$-path in $S$ (thus being a shortest path), and $P_2$ be the unique $uv_2$-path in $S$. By defining $C' := P_1 \cup P_2 \cup \{e\}$, we obtain

$$\begin{aligned} w(C) \leq w(C') \leq w(P_1) + w(P_2) + w(e) &\overset{(6)}{\leq} 2 \cdot \Delta(U) + w(e) \\ &\overset{(7)}{\leq} 4 \cdot d_{\text{HST}}(v_1, v_2) + w(e). \end{aligned}$$

□

**Theorem 6** *Every weighted graph* $(G, w)$ *with total edge weight* $W$ *admits a weakly fundamental cycle basis of length at most* $W \cdot O(\log n \log \log n)$*. Such a basis can be computed in polynomial time.*

*Proof.* We apply Algorithm 1 to construct a weakly fundamental cycle basis $\mathcal{B}$. The length $w(\mathcal{B})$ of $\mathcal{B}$ is the sum of the weights of all the basic circuits which we add in Steps 18 and 24 of the algorithm. For every non-tree edge $e = \{v_1, v_2\} \in E(G) \setminus F$ according to Lemma 4 and Lemma 5, the length $w(C)$ of the circuit $C$ that is added while processing $e$ is bounded by

$$w(C) \leq w(e) + 5 \cdot d_{\text{HST}}(v_1, v_2).$$

By summing this bound over all edges—although only non-tree edges with respect to $F$ actually count—we bound $w(\mathcal{B})$ from above with

$$\sum_{\{u,v\} \in E} w(e) + 5 \cdot d_{\text{HST}}(u, v) \overset{\text{Thm.3}}{\leq} (5 \cdot \alpha(n) + 1) \cdot W, \tag{10}$$

where $\alpha(n) \in O(\log n \log \log n)$ refers to Bartal's coefficient in Theorem 3. □

We remark that no effort was made to optimize constant factors in this proof.

## 4 Strictly Fundamental Circuits

In this section we show that for every unweighted undirected graph there exists a strictly fundamental cycle basis of size $O(n^2)$. Our proof is based on a recent result of Abraham et al. ([ABN07]) that is described below.

Given a spanning tree $T$ and a pair of vertices $x, y \in V$ as above, the *distortion* of the pair $x, y$ with respect to $T$, denoted $\text{distort}_T(x, y)$, is defined as the ratio between the distance between $x$ and $y$ in $T$ and the distance between $x$ and $y$ in $G$, i.e.,

$$\text{distort}_T(x, y) \;=\; \frac{d_T(x, y)}{d_G(x, y)} \,.$$

The $\ell_1$-*distortion* of a tree $T$ is defined as the expectation of the value of $\text{distort}_T(x, y)$, where the pair of vertices $\{x, y\}$ is drawn uniformly at random from the set of all (non-ordered) pairs of vertices $\binom{V}{2}$. In other words,

$$\text{distort}(T) \;=\; \mathbb{E}_{\mathcal{U}}(\text{distort}_T(x, y)) \,,$$

where $\mathcal{U}$ stands for the uniform distribution over $\binom{V}{2}$.

We will use the following result of Abraham et al. ([ABN07]).

**Theorem 7 (Thm. 2 in [ABN07])** *Any weighted graph $G$ with $n$ vertices contains a spanning tree $T$ with distort$(T) = O(1)$.*

Note that

$$\mathbb{E}_{\mathcal{U}}(\text{distort}_T(x, y)) = \frac{1}{\binom{n}{2}} \cdot \sum_{\{x,y\}\in\binom{V}{2}} \frac{d_T(x, y)}{d_G(x, y)} \,,$$

and thus,

$$\sum_{\{x,y\}\in\binom{V}{2}} \frac{d_T(x, y)}{d_G(x, y)} \;=\; O(n^2) \,.$$

¿From this point on we consider only unweighted graphs $G = (V, E)$. For every edge $e = \{u, v\} \in E$, $d_G(u, v) = 1$. Hence

$$\begin{aligned}\sum_{e=\{u,v\}\in E\setminus E(T)} d_T(u, v) &= \sum_{e=\{u,v\}\in E\setminus E(T)} \frac{d_T(u, v)}{d_G(u, v)} \;\le\; \sum_{\{x,y\}\in\binom{V}{2}} \frac{d_T(x, y)}{d_G(x, y)} \\ &= O(n^2) \,.\end{aligned}$$

Let $e_1 = \{u_1, v_1\}, e_2 = \{u_2, v_2\}, \ldots, e_\nu = \{u_\nu, v_\nu\}$ be the edges of $E \setminus E(T)$, ordered arbitrarily. Let $C_1, C_2, \ldots, C_\nu$ be the fundamental cycle basis with respect to $T$, with $C_i$ being the fundamental cycle of $T$ induced by the edge $e_i$, for $i \in \{1, 2, \ldots, \nu\}$. Then

$$\sum_{i=1}^{\nu} |C_i| \;=\; \sum_{i=1}^{\nu}(d_T(u_i, v_i) + 1) \;=\; \nu + \sum_{i=1}^{\nu} d_T(u_i, v_i) \;=\; O(n^2) \,.$$

**Theorem 8** *For any $n$-vertex graph there exists a cycle basis of length $O(n^2)$.*

## 5 Conclusions

After upper bounds of $O(n^2)$ (for unweighted graphs) and $O(W \cdot \log^2 n \log\log n)$ on the weights of Minimum Cycle Bases (MCB), we offer a polynomial-time algorithm that computes a weakly fundamental cycle basis of weight $O(W \cdot \log n \log\log n)$. Given the fact that there exist classes of graphs of large girth whose MCB has length $\Omega(W \cdot \log n)$, we consider our general bound fairly close to being asymptotically tight. At present we are working at upper bounds of $O(W \cdot \log n)$ for strictly fundamental cycle basis.

A strong form of what is now known as Deo's conjecture was also proposed by Deo et al. in [DKP82]:

Every unweighted graph has a strictly fundamental cycle basis of length at most $\frac{3}{2}n^2$.

Our approach and techniques seem inadequate to prove this statement.

We remark that for our construction to work it is crucial that the HST that it uses provides bounds with respect to *strong diameters*. To our knowledge the construction of HSTs due to Fakcharoenphol et al. [FRT03] that achieves the optimal $O(\log n)$ bound provides this bound only with respect to *weak* diameters. So far, we were not able to strengthen it so that the same bound will apply to strong diameters as well. Strengthening the result of [FRT03] in this way will immediately imply the tight $O(W \log n)$ bound for the MCB.

## Acknowledgment

We thank Kurt Mehlhorn for his feedback on an earlier version of the manuscript, and Alan Bertossi for making our collaboration possible.

## References

[ABN07] Ittai Abraham, Yair Bartal, and Ofer Neiman. Embedding metrics into ultrametrics and graphs into spanning trees with constant average distortion. In *Proc. of the Symp. on Discrete Algorithms, to appear*, 2007.

[AKPW95] Noga Alon, Richard M. Karp, David Peleg, and Douglas B. West. A graph-theoretic game and its application to the k-server problem. *SIAM J. Comput.*, 24(1):78–100, 1995.

[Awe85] Baruch Awerbuch. Complexity of network synchronization. *J. ACM*, 32(4):804–823, 1985.

[Bar96] Y. Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *Proceedings of the 37th IEEE FOCS*, pages 184–193, 1996.

[Bar98] Yair Bartal. On approximating arbitrary metrices by tree metrics. In *STOC*, pages 161–168, 1998.

[BK06] S. Baswana and T. Kavitha. Faster algorithms for approximate distance oracles and all-pairs small stretch paths. In *International Symposium on Foundations of Computer Science*, pages 591–602, 2006.

[Bol78] Béla Bollobás. *Extremal Graph Theory*. Academic Press, 1978.

[CLRS01] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms: Second Edition*. The MIT Press, 2001.

[CZ01] E. Cohen and U. Zwick. All-pairs small-stretch paths. *J. Algorithms*, 38(2):335–353, 2001.

[Dij59] E.W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271, 1959.

[DKP82] Narsingh Deo, M.S. Krishnomoorthy, and G.M. Prabhu. Algorithms for generating fundamental cycles in a graph. *ACM Transactions on Mathematical Software*, 8(1):26–42, 1982.

[EEST05] Michael Elkin, Yuval Emek, Daniel A. Spielman, and Shang-Hua Teng. Lower-stretch spanning trees. In Harold N. Gabow and Ronald Fagin, editors, *STOC*, pages 494–503. ACM, 2005.

[FRT03] Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. In *STOC*, pages 448–455. ACM, 2003.

[HM94] D. Hartvigsen and R. Mardon. The all-pairs min cut problem and the minimum cycle basis problem on planar graphs. *Journal of Discrete Mathematics*, 7(3):403–418, 1994.

[Hor87] Joseph Douglas Horton. A polynomial-time algorithm to find the shortest cycle basis of a graph. *SIAM Journal on Computing*, 16(2):358–366, 1987.

[KMM07] Telikepalli Kavitha, Kurt Mehlhorn, and Dimitrios Michail. New approximation algorithms for minimum cycle bases of graphs. In Wolfgang Thomas and Pascal Weil, editors, *STACS 2007*, volume 4393 of *Lecture Notes in Computer Science*, pages 512–523. Springer, 2007.

[KMMP04] Telikepalli Kavitha, Kurt Mehlhorn, Dimitrios Michail, and Katarzyna E. Paluch. A faster algorithm for minimum cycle basis of graphs. In Josep Diaz, Juhani Karhumäki, Arto Lepistö, and Donald Sanella, editors, *ICALP*, volume 3142 of *Lecture Notes in Computer Science*, pages 846–857. Springer, 2004.

[Lie03] Christian Liebchen. Finding short integral cycle bases for cyclic timetabling. In Giuseppe Di Battista and Uri Zwick, editors, *ESA*, volume 2832 of *Lecture Notes in Computer Science*, pages 715–726. Springer, 2003.

[MM07] Kurt Mehlhorn and Dimitrios Michail. Minimum cycle bases: Faster and simpler. Kurt Mehlhorn's List of Publications 196, Max-Planck-Institute Saarbrücken, 2007.

[Sey95] P. D. Seymour. Packing directed cycles fractionally. *Combinatorica*, 15(2):281–288, 1995.

[Tar75] R. E. Tarjan. Efficiency of a good but not linear set union algorithm. *J. of ACM*, 22(2):215–225, 1975.