I would like to thank Anna Lenzhen for pointing out an error in the proof of Lemma 2.11, and Matthieu Dussaule for pointing out an error in Proposition 3.3.

Here are the corrected versions.

Lemma 2.11. *Let μ be a probability distribution of finite support of diameter D. Let $X_0 \supset X_1 \supset X_2 \supset \cdots$ be a sequence of nested closed subsets of $\overline{G}$ with the following properties:*

$$1 \notin X_0 \tag{1}$$
$$X \setminus X_i \cap X_{i+1} = \varnothing \tag{2}$$
$$d(X \setminus X_i, X_{i+1}) \geqslant D \tag{3}$$

Furthermore, suppose there is a constant $0 < \epsilon < \frac{1}{2}$ such that for any $x \in X_i \setminus X_{i+1}$,

$$\nu_x(X_{i+2}) \leqslant \epsilon, \tag{4}$$
$$\nu_x(X \setminus X_{i-1}) \leqslant \epsilon, \tag{5}$$

then there are constants $c < 1$ and K, which only depend on ϵ, such that $\nu(X_i) \leqslant c^i$ and $\mu_n(X_i) \leqslant Kc^i$.

Proof. By properties (1), (2) and Proposition 2.4, any sequence of points which converges into the limit set of X_{i+2} must contain points in X_{i+1}. As the diameter of the support of μ is D, property (3) implies that any sample path which converges into X_{i+2} must contain at least one point in $X_i \setminus X_{i+1}$. Therefore, in order to find an upper bound for the probability a sample converges into X_{i+2}, we can condition on the location at which the sample path first hits $X_i \setminus X_{i+1}$. Let F be the (improper) distribution of first hitting times in X_i, i.e. $F(x)$ is equal to the probability that a sample path first hits $x \in X_i$. This is an improper distribution in general as $F(X_i) = \sum_{x \in X_i} F(x)$ may be strictly less than one, as there may be sample paths which never hit X_i. As F is supported on $X_i \setminus X_{i+1}$,

$$\nu(X_{i+2}) = \sum_{x \in X_i \setminus X_{i+1}} F(x) \nu_x(X_{i+2}).$$

For all $x \in X_i \setminus X_{i+1}$, there is an upper bound $\nu_x(X_{i+2}) \leqslant \epsilon$, by property (4), so

$$\nu(X_i) \leqslant \epsilon F(X_i). \tag{6}$$

Not all sample paths which converge to X_{i-1} need to hit X_i, but those that hit X_i and then converge to X_{i-1}, give a lower bound on $\nu(X_{i-1})$, i.e.

$$\nu(X_{i-1}) \geqslant \sum_{x \in X_i \setminus X_{i+1}} F(x) \nu_x(X_{i-1}).$$

By property (5), $\nu_x(X_{i-1}) \geqslant 1 - \epsilon$, so

$$\nu(X_{i-1}) \geqslant (1 - \epsilon) F(X_i) \tag{7}$$

Therefore, combining (6) and (7), gives

$$\frac{\nu(X_{i+2})}{\nu(X_{i-1})} \leqslant \frac{\epsilon}{1 - \epsilon} < 1,$$

as $\epsilon < \frac{1}{2}$. Therefore $\nu(X_i) \leqslant c^i$, where we may choose $c = \sqrt[3]{\epsilon/(1-\epsilon)}$.

The remaining part of the argument giving the estimate for $\mu_n(X_i)$ goes through as before. □

Lemma 3.3. *Let w_n^k be the k-iterated random walk of length n, generated by a finitely supported probability distribution μ, whose support generates a non-elementary subgroup of the mapping class group, and let $Z_i^k = 2(1 \cdot w_i^k)_{w_{i-1}^k}$. Then there are constants L, K and $c < 1$, which depend on μ but are independent of k, such that*

$$\mathbb{P}(Z_1^k + \cdots + Z_n^k \geqslant Ln) \leqslant Kc^n,$$

for all n.

Proof. We have shown that the probability that $Z_i^k \geqslant r$ decays exponentially in r, with exponential decay constants which do not depend on either k or i, or the values of any other Z_j^k for $j < i$. The Z_i^k are not independent, but Proposition 3.2 shows that

$$\mathbb{P}(Z_i^k \geqslant r \mid w_1^k, \ldots, w_{i-1}^k) \leqslant Kc^r.$$

As the Z_j^k for $j < i$ only depend on $w_1^k, \ldots w_{i-1}^k$, this implies that

$$\mathbb{P}(Z_i^k \geqslant r \mid Z_1^k, \ldots, Z_{i-1}^k) \leqslant Kc^r.$$

Therefore, the probability distribution of the sum $Z_1^k + \cdots + Z_n^k$ will be bounded above by a multiple K^n of the n-fold convolution of the exponential distribution function with itself. The rest of the proof proceeds as before. □

We remark that this is actually a standard result in the theory of stochastic dominance: if X and Y are real valued random variables, then we say that $X \lesssim Y$ if $\mathbb{P}(X \geqslant r) \leqslant \mathbb{P}(Y \geqslant r)$. If X_i and Y_i are sequences of real valued random variables with $X_i \lesssim Y_i$ for all i, and they are all independent, then it easy to show that $X_1 + \cdots + X_n \lesssim Y_1 + \cdots + Y_n$. This is not true in general if the X_i are dependent, however, if for all i, we have

$$X_i \mid X_1, \ldots, X_{i-1} \lesssim Y_i$$

then this suffices to show that $X_1 + \cdots + X_n \lesssim Y_1 + \cdots + Y_n$.