Wide Replacement Products Meet Gray Codes: Toward Optimal Small-Bias Sets

Gil Cohen Itay Cohen

Tel Aviv University

CCC 2026

Small-bias sets

A set $S \subseteq \{0,1\}^n$ is ε-biased if

$$\forall 0 \neq \tau \in \{0,1\}^n \quad |\mathbb{E}_{s\sim S}(-1)^{\langle \tau, s\rangle}| \leq \varepsilon$$

* PRGs against GF(2)-linear tests.
* Many applications in complexity (mostly via Fourier analysis).
* Equivalent to spectral expansion of Cayley graphs over the Boolean cube.
* Yield binary codes in the high-distance regime.

How small can we get?

- A random $S \subseteq \{0,1\}^n$ of size $\Theta(n/\varepsilon^2)$ is ε-biased w.h.p.
- Matches the lower bound up to a $\log(1/\varepsilon)$ factor [MRRW 1977].

How small can we get?

* A random $S \subseteq \{0,1\}^n$ of size $\Theta(n/\varepsilon^2)$ is ε-biased w.h.p.
* Matches the lower bound up to a $\log(1/\varepsilon)$ factor [MRRW 1977].

Codes

An ε-biased set $S \subseteq \{0,1\}^n$ of size m yields a binary code with rate $\rho = n/m$ and relative distance $1/2 - \varepsilon$.

Codes corresponding to set size $\Theta(n/\varepsilon^2)$ meet the Gilbert-Varshamov bound.

Explicit constructions

[Naor-Naor STOC 1990] $$\frac{n}{\varepsilon^{O(1)}}$$

[Alon-Goldreich-Hastad-Peralta FOCS 1990] $$\frac{n^2}{\varepsilon^2}$$

[BenAroya-TaShma FOCS 2009] $$\left(\frac{n}{\varepsilon^2}\right)^{5/4}$$

[TaShma STOC 2017] $$\frac{n}{\varepsilon^{2+o(1)}}$$

Explicit constructions

[Naor-Naor STOC 1990]	$\dfrac{n}{\varepsilon^{O(1)}}$	Expanders
[Alon-Goldreich-Hastad-Peralta FOCS 1990]	$\dfrac{n^2}{\varepsilon^2}$	Algebraic
[BenAroya-TaShma FOCS 2009]	$\left(\dfrac{n}{\varepsilon^2}\right)^{5/4}$	Algebraic-Geometric
[TaShma STOC 2017]	$\dfrac{n}{\varepsilon^{2+o(1)}}$	Expanders

Explicit constructions

[Naor-Naor STOC 1990] $\dfrac{n}{\varepsilon^{O(1)}}$ Expanders

Algebraic-Geometric $\dfrac{n}{\varepsilon^3}$

[Alon-Goldreich-Håstad-Peralta FOCS 1990] $\dfrac{n^2}{\varepsilon^2}$ Algebraic

[BenAroya-TaShma FOCS 2009] $\left(\dfrac{n}{\varepsilon^2}\right)^{5/4}$ Algebraic-Geometric

[TaShma STOC 2017] $\dfrac{n}{\varepsilon^{2+o(1)}}$ Expanders

Other recent approaches

[Kopparty-TaShma-Yakirevitch CCC 2026] Algebraic-Geometric

[C-Doron-Goldgraber-Manket ICALP 2026] Algebraic-Geometric

[Hsieh-Mohanty-Zhang FOCS 2026] Expanders

[Blanc-Doron CCC 2022] HDX

[TaShma STOC 2017] $\dfrac{n}{\varepsilon^{2+\alpha}}$ where $\alpha = \tilde{O}\left(\left(\dfrac{1}{\log 1/\varepsilon}\right)^{1/3}\right)$

[TaShma STOC 2017] $\frac{n}{\varepsilon^{2+\alpha}}$ where $\alpha = \tilde{O}\left(\left(\frac{1}{\log 1/\varepsilon}\right)^{1/3}\right)$

Holy grail $1/3$ => 1 would yield set size $n \cdot \tilde{O}(\varepsilon^{-2})$

[TaShma STOC 2017] $\dfrac{n}{\varepsilon^{2+\alpha}}$ where $\alpha = \tilde{O}\left(\left(\dfrac{1}{\log 1/\varepsilon}\right)^{1/3}\right)$

Holy grail $1/3$ => 1 would yield set size $n \cdot \tilde{O}(\varepsilon^{-2})$

Our quantitative result $1/3$ => $1/2$

[TaShma STOC 2017] $\frac{n}{\varepsilon^{2+\alpha}}$ where $\alpha = \tilde{O}\left(\left(\frac{1}{\log 1/\varepsilon}\right)^{1/3}\right)$

Holy grail $1/3$ => 1 would yield set size $n \cdot \tilde{O}(\varepsilon^{-2})$

Our quantitative result $1/3$ => $1/2$

Our more significant contribution

Overcoming one of the two barriers toward the holy grail.

Bias Reduction

Observation. If S is an ε_0-biased set then

$$S + S = \{s_1 + s_2 : s_1, s_2 \in S\}$$

is ε_0^2-biased.

However, this construction has size $|S|^2 > n^2$.

Observation. If S is an ε_0-biased set then

$$S + S = \{s_1 + s_2 : s_1, s_2 \in S\}$$

is ε_0^2-biased.

However, this construction has size $|S|^2 > n^2$.

Idea [Rozenman-Wigderson]. Derandomize using an expander:

$$S +_G S = \{s_1 + s_2 : s_1 \sim_G s_2\}$$

Claim. This is an $\varepsilon_0^2 + \omega$- biased set.

Claim. $S +_G S = \{s_1 + s_2 : s_1 \sim_G s_2\}$ is an $\varepsilon_0^2 + \omega$- biased set.

Claim. $S +_G S = \{s_1 + s_2 : s_1 \sim_G s_2\}$ is an $\varepsilon_0^2 + \omega$- biased set.

Proof.

A given test τ induces a ± 1 labeling on the vertices of G: vertex s is labeled by $(-1)^{\langle s,\tau \rangle}$. Encode these labels as the diagonal of Π. Then,

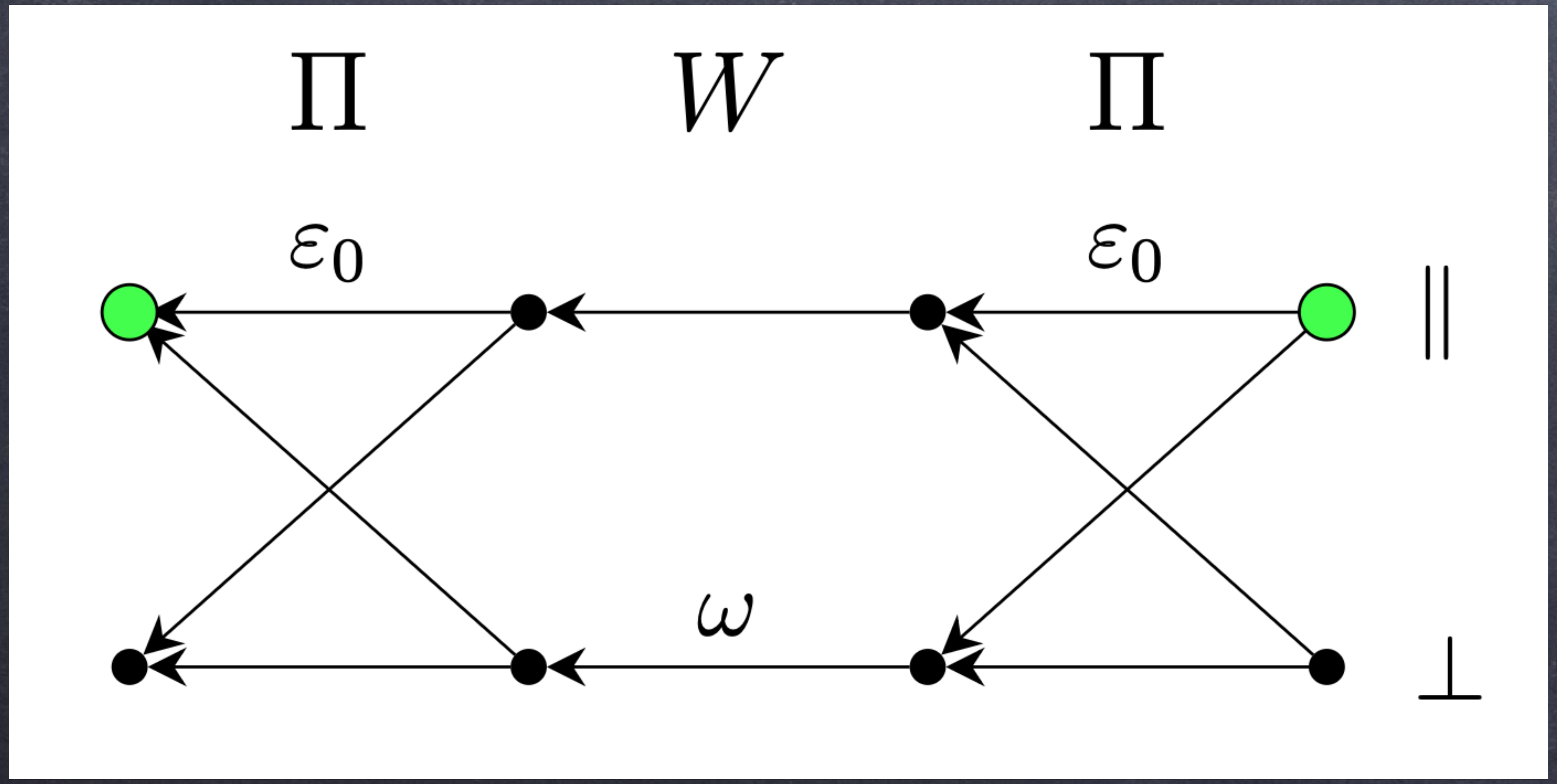
$$\varepsilon_\tau = \left| \mathbf{1}^T \Pi W \Pi \mathbf{1} \right|$$

Claim. $S +_G S = \{s_1 + s_2 : s_1 \sim_G s_2\}$ is an $\varepsilon_0^2 + \omega$- biased set.

Proof.

A given test τ induces a ± 1 labeling on the vertices of G: vertex s is labeled by $(-1)^{\langle s,\tau\rangle}$. Encode these labels as the diagonal of Π. Then,

$$\varepsilon_\tau = \left|\mathbf{1}^T \Pi W \Pi \mathbf{1}\right|$$



$$\Rightarrow \qquad \varepsilon_\tau \leq \varepsilon_0^2 + \omega$$

Longer walks - dream version

2-vertex walks yield bias $\varepsilon_0^2 + \omega$.

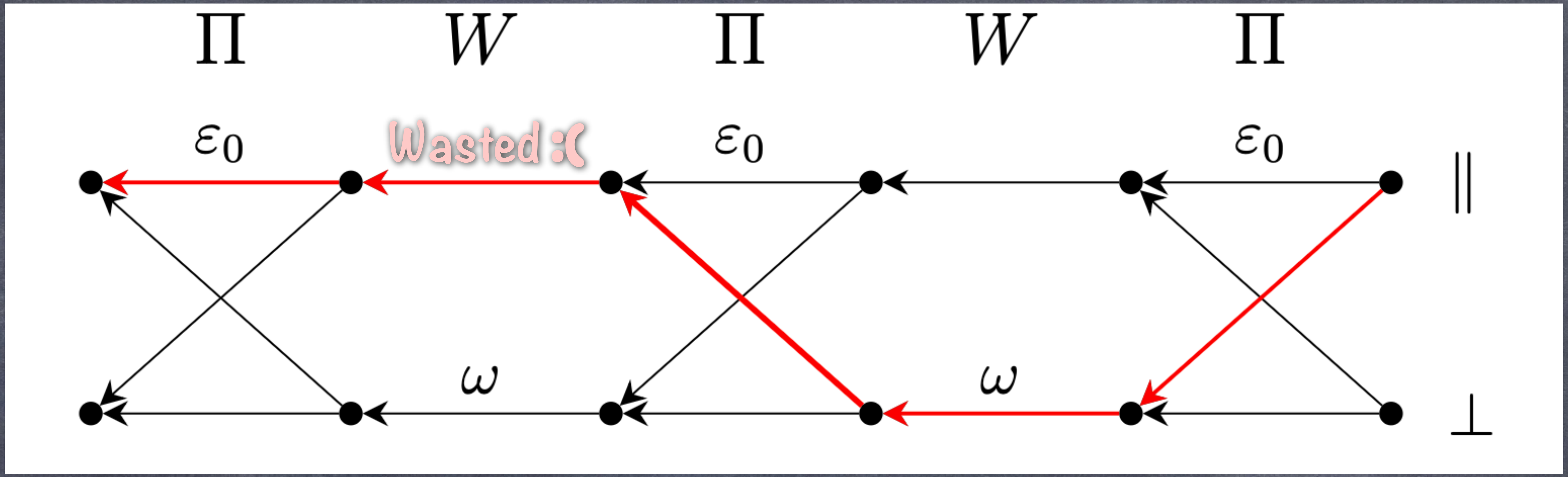
If t-vertex walks give bias $\varepsilon_0^t + \omega^t$ (or $\omega^{t-O(1)}$) then, using a Ramanujan graph for G, starting with $\varepsilon_0 \approx O(1)$, and taking $t = O(\log 1/\varepsilon)$ yields an ε-bias set of size

$$n \cdot \tilde{O}(\varepsilon^{-2})$$

The factor 2 comes from the $\sqrt{\ }$ in the Ramanujan bound $\omega \approx 2/\sqrt{d}$.

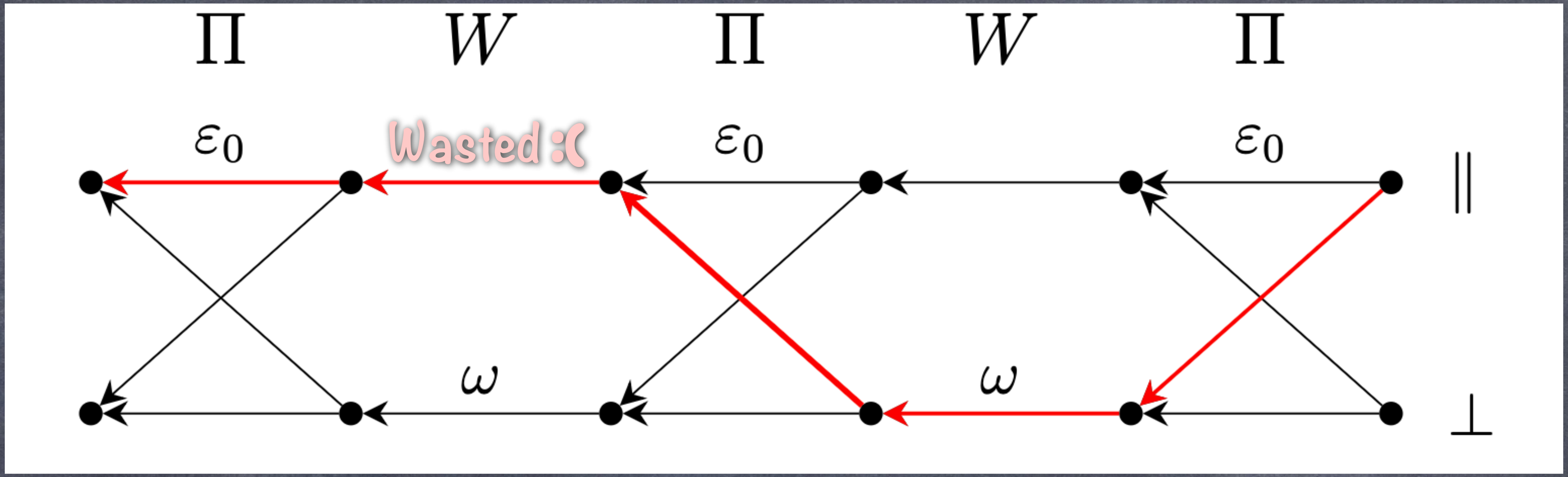
Back to reality

$$\varepsilon_\tau = \left| \mathbf{1}^T \Pi W \Pi W \Pi \mathbf{1} \right|$$



Back to reality

$$\varepsilon_\tau = \left| \mathbf{1}^T \Pi W \Pi W \Pi \mathbf{1} \right|$$



Informally, every other step may be wasted, so the bound is roughly

$$\varepsilon_0^t + \omega^{t/2}$$

This 2-loss x the Ramanujan 2-“loss” yields set size $n \cdot \tilde{O}(\varepsilon^{-4})$.

TaShma's Idea

Protect your walk from the adversary

Use a fresh register at each of the t steps in a way that guarantees bias

$$\varepsilon_0^t + \omega^{t-O(1)}$$

Based on [BenAroya-TaShma STOC 2008]'s wide-replacement product.

Since the set size increases exponentially in t, optimizing the tradeoff yields TaShma's bound.

Our Contribution

Reuse registers

By reusing registers in a carefully chosen order, we can take length-t walks using only $O(\log t)$ registers, while still achieving $\varepsilon_0^t + \omega^{t-O(1)}$.

Since the register overhead now increases only polynomially in t, optimizing the tradeoff gives our improved bound.

Reuse registers

By reusing registers in a carefully chosen order, we can take length-t walks using only $O(\log t)$ registers, while still achieving $\varepsilon_0^t + \omega^{t-O(1)}$.

Since the register overhead now increases only polynomially in t, optimizing the tradeoff gives our improved bound.

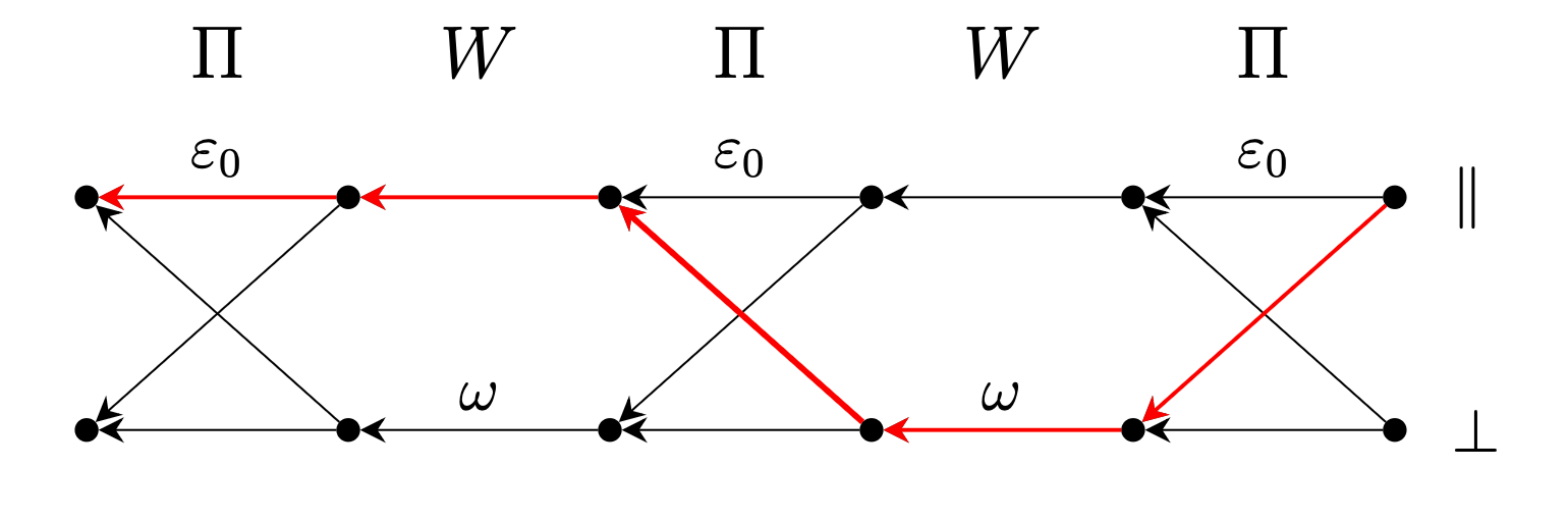
We use a Gray-code-style reuse schedule

1 2 3 1

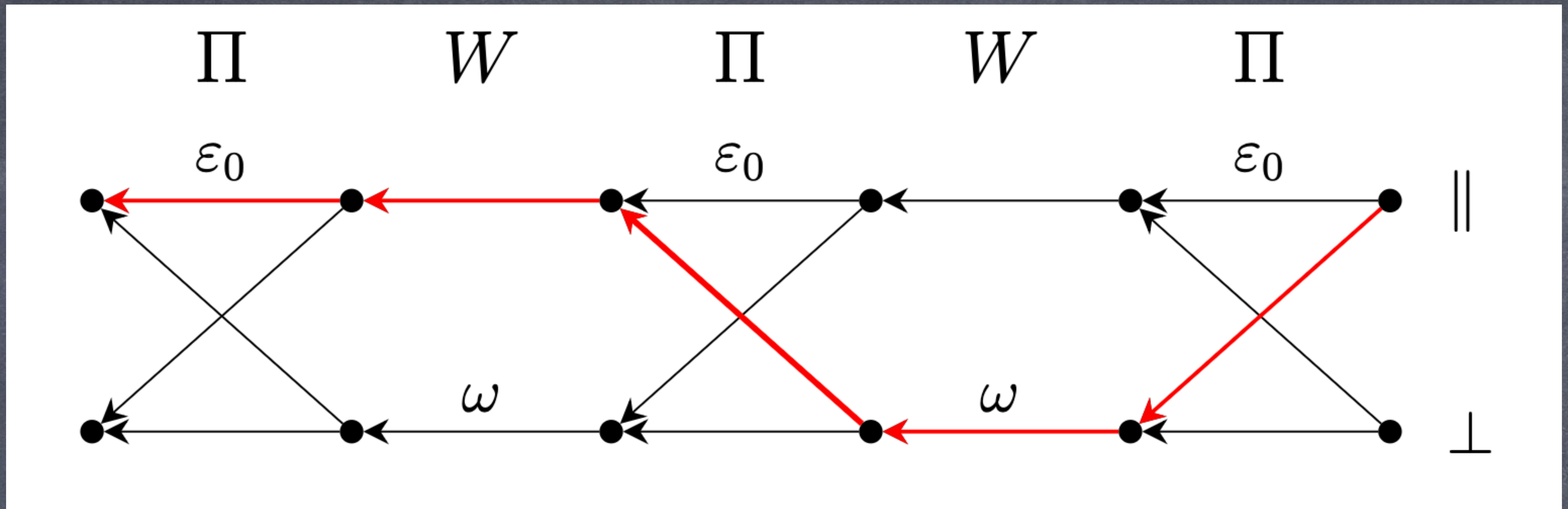
1 2 3 1 4 5 1 2 3 1

1 2 3 1 4 5 1 2 3 1 6 7 1 2 3 1 4 5 1 2 3 1

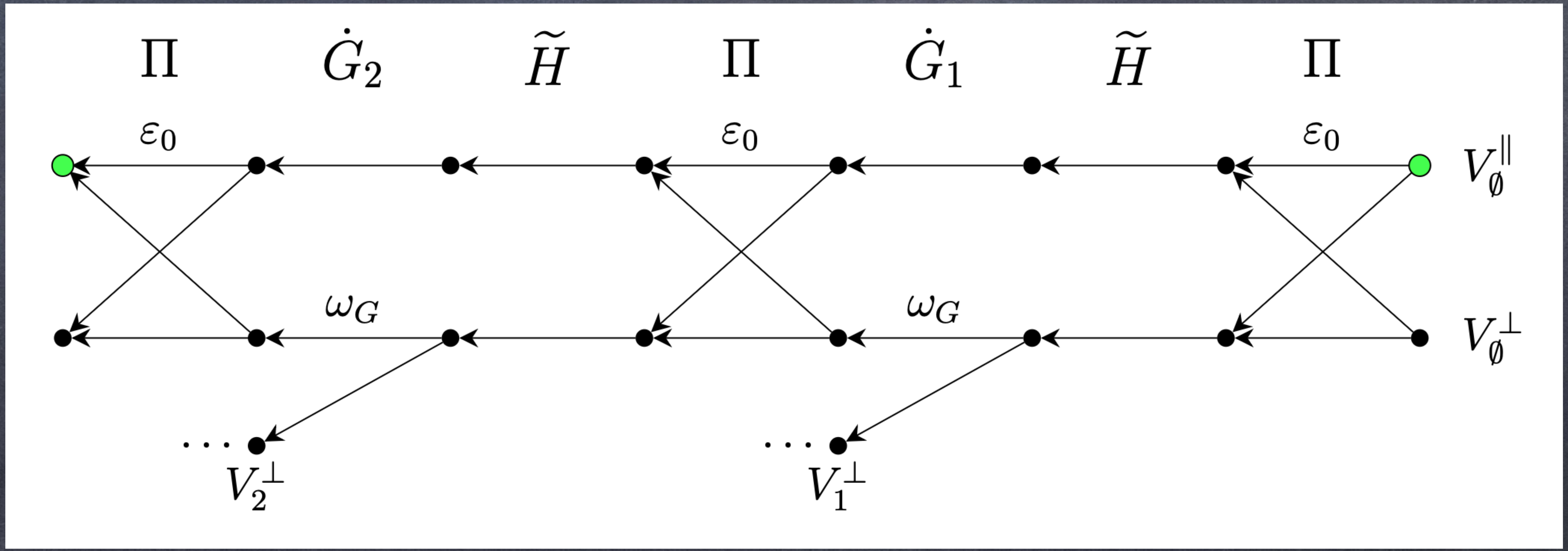
Standard walks

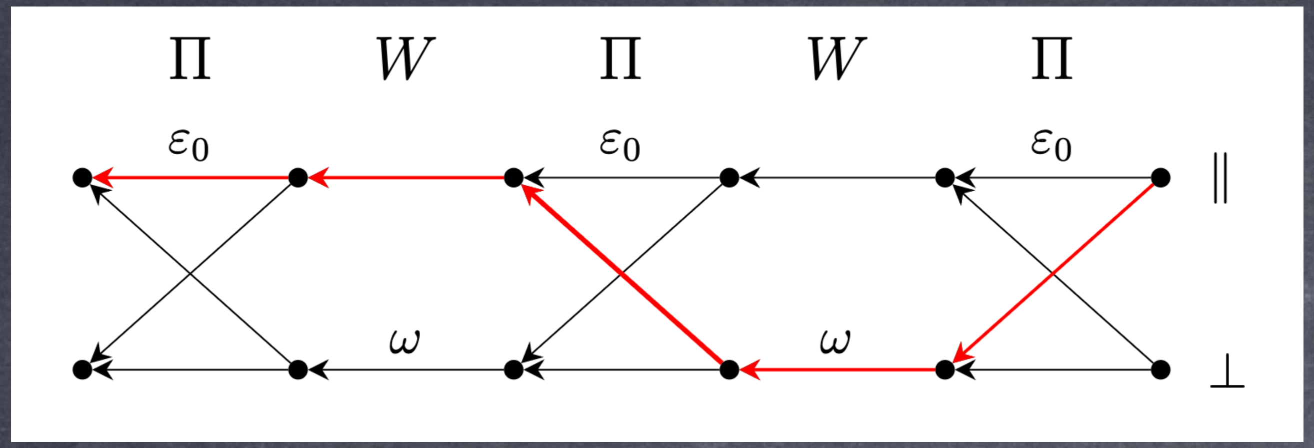


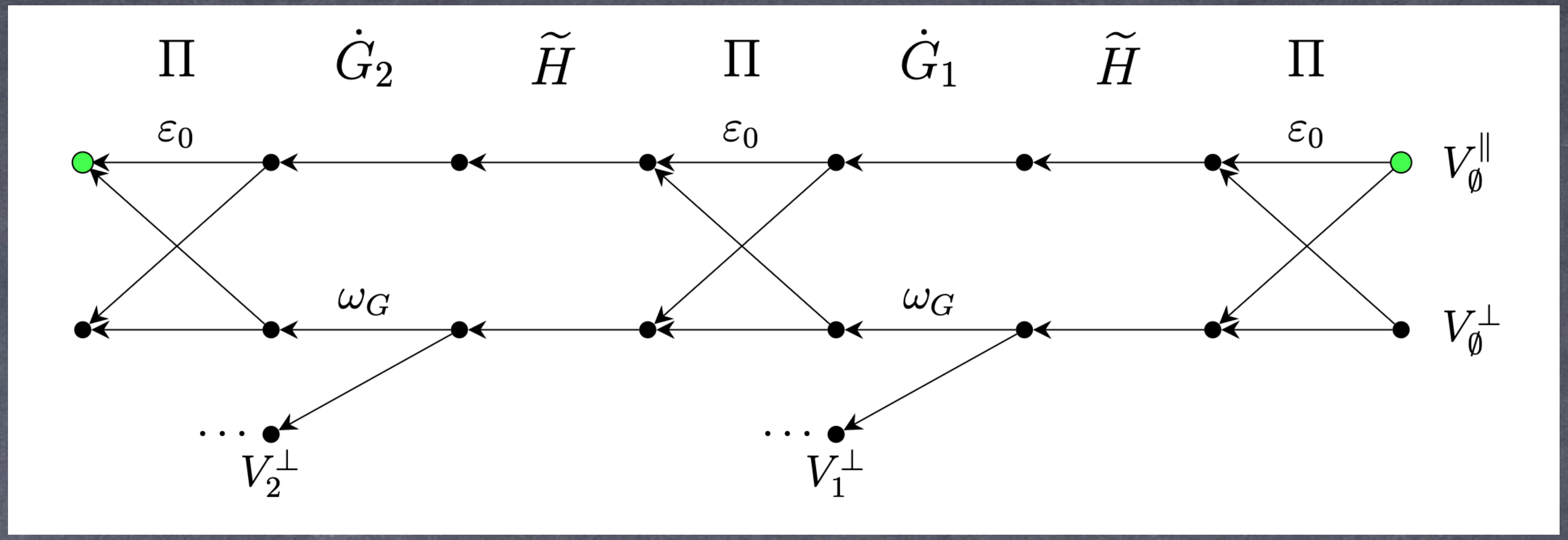
Standard walks



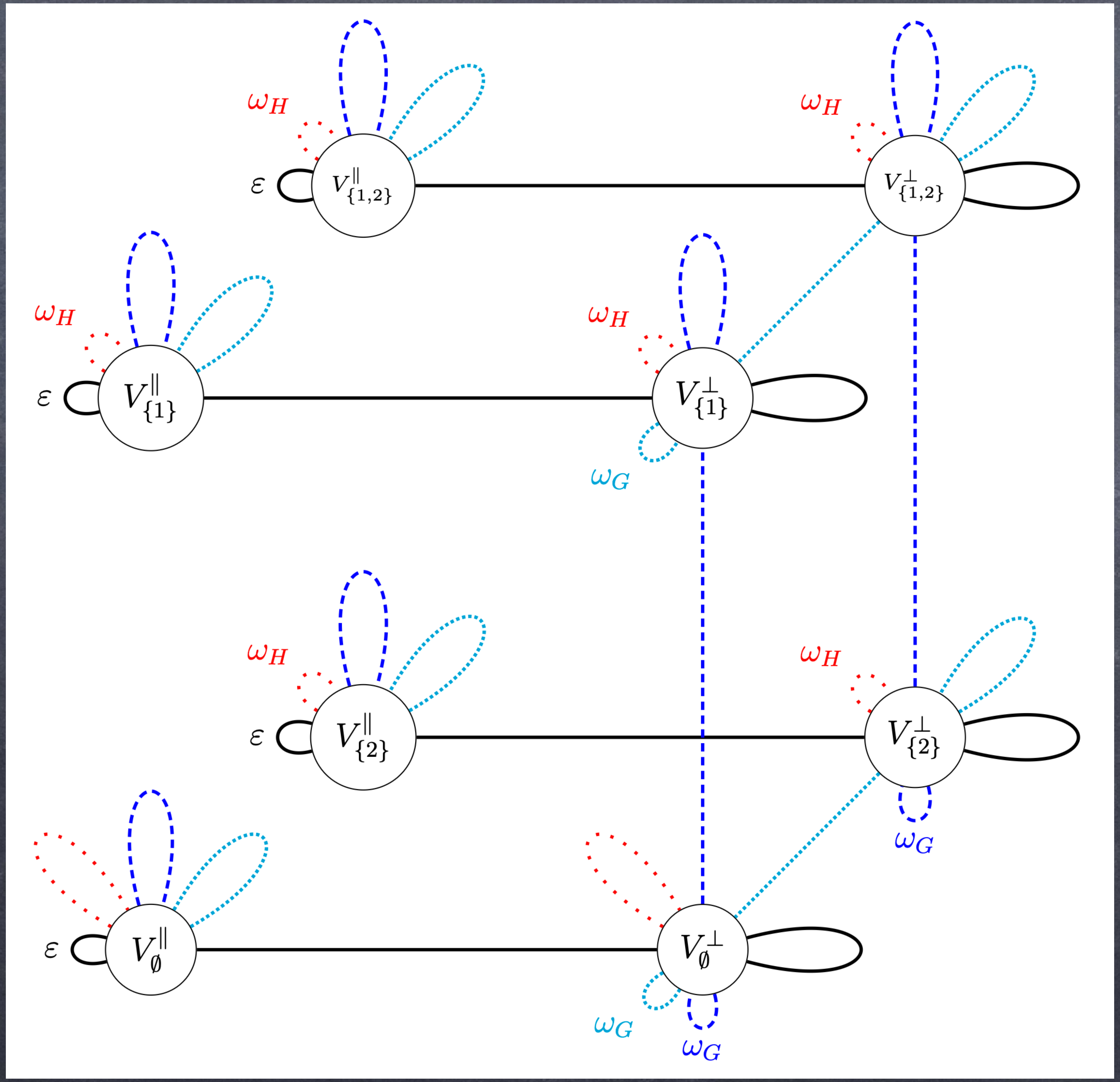
Wide walks;
fresh registers







Wide walks; reusing registers



The Second Barrier

The remaining barrier

The second barrier stems from the 2 in $\omega \approx 2/\sqrt{d}$.

Taking t steps yields an exponential blowup, 2^t.

The remaining barrier

The second barrier stems from the 2 in $\omega \approx 2/\sqrt{d}$.

Taking t steps yields an exponential blowup, 2^t.

Ways around this barrier

- Directed graphs can have spectral radius $\omega \approx 1/\sqrt{d}$
- Non-backtracking walks on undirected graphs have a linear blowup in t.
- Same for random walks on rotating expanders [C-Maor STOC 2023].

The challenge is to combine these with our framework.

Open problems

- Construct explicit ε-biased sets $S \subset \{0,1\}^n$ of size $n \cdot \tilde{O}(\varepsilon^{-2})$.
- Which bound is tight—the known lower bound or the probabilistic upper bound?
- Can our register-reuse technique be applied elsewhere?

Thanks!