THE STATISTICAL RESTRICTED ISOMETRY PROPERTY AND 
THE WIGNER SEMICIRCLE DISTRIBUTION OF INCOHERENT 
DICTIONARIES

SHAMGAR GUREVICH AND RONNY HADANI

ABSTRACT. In this paper we formulate and prove a statistical version of the 
Candès-Tao restricted isometry property (SRIP for short) which holds in gen-
eral for any incoherent dictionary which is a disjoint union of orthonormal 
bases. In addition, we prove that, under appropriate normalization, the eigen-
values of the associated Gram matrix fluctuate around \( \lambda = 1 \) according to the 
Wigner semicircle distribution. The result is then applied to various dictio-
naries that arise naturally in the setting of finite harmonic analysis, giving, in 
particular, a better understanding on a remark of Applebaum-Howard-Searle-
Calderbank concerning RIP for the Heisenberg dictionary of chirp like func-
tions.

0. Introduction

Digital signals, or simply signals, can be thought of as complex valued functions 
on the finite field \( \mathbb{F}_p \), where \( p \) is a prime number. The space of signals \( \mathcal{H} = \mathbb{C}(\mathbb{F}_p) \) 
is a Hilbert space of dimension \( p \), with the inner product given by the standard 
formula

\[
\langle f, g \rangle = \sum_{t \in \mathbb{F}_p} f(t) \overline{g(t)}.
\]

A dictionary \( \mathcal{D} \) is simply a set of vectors (also called atoms) in \( \mathcal{H} \). The number 
of vectors in \( \mathcal{D} \) can exceed the dimension of the Hilbert space \( \mathcal{H} \), in fact, the most 
interesting situation is when \( |\mathcal{D}| > p = \dim \mathcal{H} \). In this set-up we define a resolution 
of the Hilbert space \( \mathcal{H} \) via \( \mathcal{D} \), which is the morphism of vector spaces

\[ \Theta : \mathbb{C}(\mathcal{D}) \to \mathcal{H}, \]

given by \( \Theta(f) = \sum_{\varphi \in \mathcal{D}} f(\varphi) \varphi \), for every \( f \in \mathbb{C}(\mathcal{D}) \). A more concrete way to think 
of the morphism \( \Theta \) is as a \( p \times |\mathcal{D}| \) matrix with the columns being the atoms in \( \mathcal{D} \).

In the last two decades [13], and in particular in recent years [5, 6, 7, 8, 9, 10], 
resolutions of Hilbert spaces became an important tool in signal processing, in 
particular in the emerging theories of sparsity and compressive sensing.

0.1. The statistical restricted isometry property. A useful property of a reso-
novation is the restricted isometry property (RIP for short) defined by Candès-Tao 
in [9]. Fix a natural number \( n \in \mathbb{N} \) and a pair of positive real numbers \( \delta_1, \delta_2 \in \mathbb{R}_{>0} \).

Definition 0.1. A dictionary \( \mathcal{D} \) satisfies the restricted isometry property with co-
efficients \( (\delta_1, \delta_2, n) \) if for every subset \( S \subset \mathcal{D} \) such that \( |S| \leq n \) we have

\[
(1 - \delta_2) \|f\| \leq \|\Theta(f)\| \leq (1 + \delta_1) \|f\|,
\]
for every function $f \in C(\mathcal{D})$ which is supported on the set $S$.

Equivalently, RIP can be formulated in terms of the spectral radius of the corresponding Gram operator. Let $G(S)$ denote the composition $\Theta_S^* \circ \Theta_S$ with $\Theta_S$ denoting the restriction of $\Theta$ to the subspace $C_S(\mathcal{D}) \subset C(\mathcal{D})$ of functions supported on the set $S$. The dictionary $\mathcal{D}$ satisfies $(\delta_1, \delta_2, n)$-RIP if for every subset $S \subset D$ such that $|S| \leq n$ we have

$$\delta_2 \leq \|G(S) - Id_S\| \leq \delta_1,$$

where $Id_S$ is the identity operator on $C_S(\mathcal{D})$.

It is known [4, 10] that the RIP holds for random dictionaries. However, one would like to address the following problem [2, 12, 11, 21, 22, 23, 24, 26, 25, 28, 29]:

**Problem 0.2.** Find deterministic construction of a dictionary $\mathcal{D}$ with $|\mathcal{D}| \gg p$ which satisfies RIP with coefficients in the critical regime

$$\delta_1, \delta_2 \ll 1 \text{ and } n = \alpha \cdot p,$$

for some constant $0 < \alpha < 1$.

0.2. Incoherent dictionaries. Fix a positive real number $\mu \in \mathbb{R}_{>0}$. The following notion was introduced in [11, 14] and was used to study similar problems in [28, 29]:

**Definition 0.3.** A dictionary $\mathcal{D}$ is called incoherent with coherence coefficient $\mu$ (also called $\mu$-coherent) if for every pair of distinct atoms $\varphi, \phi \in \mathcal{D}$

$$|\langle \varphi, \phi \rangle| \leq \frac{\mu}{\sqrt{p}}.$$

In this paper we will explore a general relation between RIP and incoherence. Our motivation comes from three examples of incoherent dictionaries which arise naturally in the setting of finite harmonic analysis (for the sake of completeness we review the construction of these examples in Section 3):

- The first example [19, 20], referred to as the Heisenberg dictionary $\mathcal{D}_H$, is constructed using the Heisenberg representation of the finite Heisenberg group $H(\mathbb{F}_p)$. The Heisenberg dictionary is of size approximately $p^2$ and its coherence coefficient is $\mu = 1$.

- The second example [16, 18], which is referred to as the oscillator dictionary $\mathcal{D}_O$, is constructed using the Weil representation of the finite symplectic group $SL_2(\mathbb{F}_p)$. The oscillator dictionary is of size approximately $p^3$ and its coherence coefficient is $\mu = 4$.

- The third example [16, 18], referred to as the extended oscillator dictionary $\mathcal{D}_{EO}$, is constructed using the Heisenberg-Weil representation of the finite Jacobi group $J(\mathbb{F}_p) = SL_2(\mathbb{F}_p) \rtimes H(\mathbb{F}_p)$. The extended oscillator dictionary is of size approximately $p^5$ and its coherence coefficient is $\mu = 4$.

The three examples of dictionaries we just described constitute reasonable candidates for solving Problem 0.2. They are large in the sense that $|\mathcal{D}| \gg p$, and empirical evidences suggest (see [2] for the case of $\mathcal{D}_H$) that they might satisfy RIP with coefficients in the critical regime (0.1). We summarize this as follows:

**Question:** Do the dictionaries $\mathcal{D}_H, \mathcal{D}_O$ and $\mathcal{D}_{EO}$ satisfy the RIP with coefficients $\delta_1, \delta_2 \ll 1$ and $n = \alpha \cdot p$, for some $0 < \alpha < 1$?
0.3. Main results. In this paper we formulate a relaxed statistical version of RIP, called statistical isometry property (SRIP for short) and we prove that it holds for any incoherent dictionary \( \mathcal{D} \) which is, in addition, a disjoint union of orthonormal bases:

\[
\mathcal{D} = \bigsqcup_{x \in \mathcal{X}} B_x,
\]

where \( B_x = \{ b_x^1, ..., b_x^p_x \} \) is an orthonormal basis of \( \mathcal{H} \), for every \( x \in \mathcal{X} \).

0.3.1. The statistical restricted isometry property. Let \( \mathcal{D} \) be an incoherent dictionary of the form (0.2). Roughly, the statement is that for \( S \subset \mathcal{D} \), \( |S| = n \) with \( n = p^{1-\varepsilon} \), for \( 0 < \varepsilon < 1 \), chosen uniformly at random, the operator norm \( \| G(S) - \text{Id}_S \| \) is small with high probability.

Theorem 0.4 (SRIP property). For every \( k \in \mathbb{N} \), there exists a constant \( C(k) \) such that the probability

\[
\Pr \left( \| G(S) - \text{Id}_S \| \geq p^{1-\varepsilon} / 2 \right) \leq C(k) p^{1-\varepsilon k / 2}.
\]

The above theorem, in particular, implies that the probability \( \Pr \left( \| G(S) - \text{Id}_S \| \geq p^{1-\varepsilon} / 2 \right) \to 0 \) as \( p \to \infty \) faster then \( p^{-l} \) for any \( l \in \mathbb{N} \).

0.3.2. The statistics of the eigenvalues. A natural thing to know is how the eigenvalues of the Gram operator \( G(S) \) fluctuate around 1. In this regard, we study the spectral statistics of the normalized error term

\[
E(S) = (p/n)^{1/2} (G(S) - \text{Id}_S).
\]

Let \( \rho_{E(S)} = \sum_{i=1}^{n} \delta_{\lambda_i} \) denote the spectral distribution of \( E(S) \) where \( \lambda_i \), \( i = 1, ..., n \), are the real eigenvalues of the Hermitian operator \( E(S) \). We prove that \( \rho_{E(S)} \) converges in probability as \( p \to \infty \) to the Wigner semicircle distribution

\[
\rho_{SC}(x) = \frac{1}{\pi} \sqrt{4 - x^2} \cdot 1_{[-2, 2]}(x)
\]

where \( 1_{[-2, 2]} \) is the characteristic function of the interval \([-2, 2]\).

Theorem 0.5 (Semicircle distribution). We have

\[
\lim_{p \to \infty} \rho_{E(S)}^p = \rho_{SC}.
\]

Remark 0.6. A limit of the form (0.4) is familiar in random matrix theory as the asymptotic of the spectral distribution of Wigner matrices. Interestingly, the same asymptotic distribution appears in our situation, albeit, the probability spaces are of a different nature (our probability spaces are, in particular, much smaller).

In particular, Theorems 0.4, 0.5 can be applied to the three examples \( \mathcal{D}_H, \mathcal{D}_O \) and \( \mathcal{D}_{EO} \), which are all of the appropriate form (0.2). Finally, our result gives new information on a remark of Applebaum-Howard-Searle-Calderbank [2] concerning RIP of the Heisenberg dictionary.

Remark 0.7. For practical applications, it might be important to compute explicitly the constants \( C(k) \) which appears in (0.3). This constant depends on the incoherence coefficient \( \mu \), therefore, for a fixed \( p \), having \( \mu \) as small as possible is preferable.
0.3.3. Structure of the paper. The paper consists of four sections except of the introduction.

In Section 1, we develop the statistical theory of systems of incoherent orthonormal bases. We begin by specifying the basic set-up. Then we proceed to formulate and prove the main Theorems of this paper - Theorem 1.2, Theorem 1.3, and Theorem 1.4. The main technical statement underlying the proofs is formulated in Theorem 1.5. In Section 2, we prove Theorem 1.5. In Section 3, we review the constructions of the dictionaries $D_H$, $D_O$, and $D_{EO}$. Finally, in Appendix A, we prove all technical statements which appear in the body of the paper.

Acknowledgement 0.8. It is a pleasure to thank our teacher J. Bernstein for his continuos support. We are grateful to N. Sochen for many stimulating discussions. We thank F. Bruckstein, R. Calderbank, M. Elad, Y. Eldar, R. Kimmel, and A. Sahai for sharing with us some of their thoughts about signal processing. We are grateful to R. Howe, A. Man, M. Revzen and Y. Zak for explaining us the notion of mutually unbiased bases.

1. The statistical theory of incoherent bases

1.1. Standard Terminology.

1.1.1. Terminology from asymptotic analysis. Let $\{a_p\}, \{b_p\}$ be a pair of sequences of positive real numbers. We write $a_p = O(b_p)$ if there exists $C > 0$ and $P_0 \in \mathbb{N}$ such that $a_p \leq C \cdot b_p$ for every $p \geq P_0$. We write $a_p = o(b_p)$ if $\lim_{p \to \infty} a_p/b_p = 0$. Finally, we write $a_p \sim b_p$ if $\lim_{p \to \infty} a_p/b_p = 1$.

1.1.2. Terminology from set theory. Let $n \in \mathbb{N}_{\geq 1}$. We denote by $[1, n]$ the set $\{1, 2, \ldots, n\}$. Given a finite set $A$, we denote by $|A|$ the number of elements in $A$.

1.2. Basic set-up.

1.2.1. Incoherent orthonormal bases. Let $\{(\mathcal{H}_p, \langle -,-\rangle_p)\}$ be a sequence of Hilbert spaces such that $\dim \mathcal{H}_p = p$.

Definition 1.1. Two (sequences of) orthonormal bases $B_p, B'_p$ of $\mathcal{H}_p$ are called $\mu$-coherent if

$$|\langle b, b' \rangle| \leq \frac{\mu}{\sqrt{p}}$$

for every $b \in B_p$ and $b' \in B'_p$, and $\mu$ is some fixed (does not depend on $p$) positive real number.

Fix $\mu \in \mathbb{R}^{>0}$. Let $\{\mathcal{X}_p\}$ be a sequence of sets such that $\lim_{p \to \infty} |\mathcal{X}_p| = \infty$ (usually we will have that $p = o(|\mathcal{X}_p|)$) such that each $\mathcal{X}_p$ parametrizes orthonormal bases of $\mathcal{H}_p$ which are $\mu$-coherent pairwise., that is, for every $x \in \mathcal{X}_p$, there is an orthonormal basis $B_x = \{b^x_1, \ldots, b^x_p\}$ of $\mathcal{H}_p$ so that

$$|\langle b^x_i, b^y_j \rangle| \leq \frac{\mu}{\sqrt{p}},$$

for every $x \neq y \in \mathcal{X}_p$. Denote

$$\mathcal{D}_p = \bigsqcup_{x \in \mathcal{X}_p} B_x.$$

The set $\mathcal{D}_p$ will be referred to as incoherent dictionary or sometime more precisely as $\mu$-coherent dictionary.
1.2.2. Resolutions of Hilbert spaces. Let $\Theta_p: \mathbb{C}(\mathcal{D}_p) \to \mathcal{H}_p$ be the morphism of vector spaces given by

$$\Theta_p(f) = \sum_{b \in \mathcal{D}_p} f(b) b.$$ 

The map $\Theta_p$ will be referred to as a resolution of $\mathcal{H}_p$ via $\mathcal{D}_p$.

Convention: For the sake of clarity we will usually omit the subscript $p$ from the notations.

1.3. Statistical restricted isometry property (SRIP). The main statement of this paper concerns a formulation of a statistical restricted isometry property (SRIP for short) of the resolution maps $\Theta$.

Let $n = n(p) = p^{1-\varepsilon}$, for some $0 < \varepsilon < 1$. Let $\Omega_n = \Omega([1,n])$ denote the set of injective maps $\Omega_n = \{ S: [1,n] \hookrightarrow \mathcal{D} \}$. We consider the set $\Omega_n$ as a probability space equipped with the uniform probability measure.

Given a map $S \in \Omega_n$, it induces a morphism of vector spaces $S: \mathbb{C}([1,n]) \to \mathbb{C}(\mathcal{D})$ given by $S(\delta_i) = \delta_{S(i)}$. Let us denote by $\Theta_S: \mathbb{C}([1,n]) \to \mathcal{H}$ the composition $\Theta \circ S$ and by $G(S) \in \text{Mat}_{n \times n}(\mathbb{C})$ the Hermitian matrix

$$G(S) = \Theta_S^* \circ \Theta_S.$$ 

Concretely, $G(S)$ is the matrix $(g_{ij})$ where $g_{ij} = \langle S(i), S(j) \rangle$. In plain language, $G(S)$ is the Gram matrix associated with the ordered set of vectors $(S(1), ..., S(n))$ in $\mathcal{H}$.

We consider $G: \Omega_n \to \text{Mat}_{n \times n}(\mathbb{C})$ as a matrix valued random variable on the probability space $\Omega_n$. The following theorem asserts that with high probability the matrix $G$ is close to the unit matrix $I_n \in \text{Mat}_{n \times n}(\mathbb{C})$.

**Theorem 1.2.** Let $0 \leq \varepsilon \ll 1$ and let $k \in \mathbb{N}$ be an even number such that $ek \gg 1$

$$\Pr \left( \|G - I_n\| \geq \left(\frac{n}{p}\right)^{1/(2+\varepsilon)} \right) = O \left(\left(\frac{n}{p}\right)^{ek/(2+\varepsilon)} \cdot n \right).$$

For a proof, see Subsection 1.6.

In the above theorem, substituting $n = p^{1-\varepsilon}$ yields

$$\Pr \left( \|G - I_n\| \geq p^{-\varepsilon/(2+\varepsilon)} \right) = O \left(p^{-\varepsilon(k+1)/(2+\varepsilon)+1} \right).$$

Equivalently, Theorem 1.2 can be formulated as a statistical restricted isometry property of the resolution morphism $\Theta$.

A given $S \in \Omega_n$ defines a morphism of vector spaces $\Theta_S = \Theta \circ S: \mathbb{C}([1,n]) \to \mathcal{H}$ - in this respect, $\Theta$ can be considered as a random variable

$$\Theta: \Omega_n \to \text{Mor} (\mathbb{C}([1,n]), \mathcal{H}).$$

**Theorem 1.3** (SRIP property). Let $0 \leq \varepsilon \ll 1$ and let $k \in \mathbb{N}$ be an even number such that $ek \gg 1$

$$\Pr \left( \text{Sup}\{\|\Theta(f)\| - \|f\|\} \geq \left(\frac{n}{p}\right)^{1/(2+\varepsilon)} \right) = O \left(\left(\frac{n}{p}\right)^{ek/(2+\varepsilon)} \cdot n \right).$$
1.4. **Statistics of the error term.** Let $E$ denote the normalized error term

$$E = (p/n)^{1/2} (G - I_n).$$

Our goal is to describe the statistics of the random variable $E$. Let $\rho_{E}$ denote the spectral distribution of $E$, namely

$$\rho_{E} = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_{i}(E)},$$

where $\lambda_{1}(E) \geq \lambda_{2}(E) \geq \ldots \geq \lambda_{n}(E)$ are the eigenvalues of $E$ indexed in decreasing order (we note that the eigenvalues of $E$ are real since it is an Hermitian matrix). The following theorem asserts that the spectral distribution $\rho_{E}$ converges in probability to the Wigner semicircle distribution 

$$\rho_{SC}(x) = \frac{1}{2\pi} \sqrt{4 - x^2} \cdot 1_{[2,-2]}(x).$$

**Theorem 1.4.**

$$\lim_{p \to \infty} \rho_{E} \Pr = \rho_{SC}.$$

For a proof, see Subsection 1.7.

1.5. **The method of moments.** The proofs of Theorems 1.2, 1.4 will be based on the method of moments.

Let $m_{k}$ denote the $k$th moment of the distribution $\rho_{E}$, that is

$$m_{k} = \int x^{k} \rho_{E}(x) = \frac{1}{n} \sum_{i=1}^{n} \lambda_{i}(E)^{k} = \frac{1}{n} \text{Tr}(E^{k}).$$

Similarly, let $m_{SC,k}$ denote the $k$th moment of the semicircle distribution.

**Theorem 1.5.** For every $k \in \mathbb{N}$,

$$\lim_{p \to \infty} E(m_{k}) = m_{SC,k}. $$

In addition,

$$Var(m_{k}) = O(n^{-1}).$$

For a proof, see Section 2.

1.6. **Proof of Theorem 1.2.** Theorem 1.2 follows from Theorem 1.4 using the Markov inequality.

Let $\delta > 0$ and $k \in \mathbb{N}$ an even number. First, observe that the condition $\|G - I_n\| \geq \delta$ is equivalent to the condition $\|E\| \geq (p/n)^{1/2} \delta$ which, in turns, is equivalent to the spectral condition $\lambda_{\max}(E) \geq (p/n)^{1/2} \delta$.

Since, $\lambda_{\max}(E)^{k} \leq \lambda_{1}(E)^{k} + \lambda_{2}(E)^{k} + \ldots + \lambda_{n}(E)^{k}$ we can write

$$\Pr \left( \lambda_{\max}(E) \geq (p/n)^{1/2} \delta \right) = \Pr \left( \lambda_{\max}(E)^{k} \geq (p/n)^{k/2} \delta^{k} \right) \leq \Pr \left( \sum_{i=1}^{n} \lambda_{i}(E)^{k} \geq (p/n)^{k/2} \delta^{k} \right) = \Pr \left( m_{k} \geq n^{-1} (p/n)^{k/2} \delta^{k} \right).$$
By the triangle inequality $\mathbf{m}_k \leq |\mathbf{m}_k - E\mathbf{m}_k| + E\mathbf{m}_k$ (recall that $k$ is even, hence $\mathbf{m}_k \geq 0$) therefore we can write

$$\Pr \left( \mathbf{m}_k \geq n \left( \frac{p}{n} \right)^{k/2} \delta_k \right) \leq \Pr \left( |\mathbf{m}_k - E\mathbf{m}_k| \geq n \left( \frac{p}{n} \right)^{k/2} \delta_k - E\mathbf{m}_k \right).$$

By [1.3], $E\mathbf{m}_k = O(1)$, in addition, substituting $\delta = (n/p)^{1/(e+2)} = (p/n)^{-1/(e+2)}$ with $0 < e < 1$, we get $n^{-1} (p/n)^{k} \delta^k = n^{-1} (p/n)^{ek/2(2+e) \delta}$. Altogether, we can summarize the previous development with the following inequality

$$\Pr \left( \| \mathbf{G} - I_n \| \geq (n/p)^{1/(e+2)} \right) \leq \Pr \left( |\mathbf{m}_k - E\mathbf{m}_k| \geq n^{-1} (p/n)^{ek/2(2+e) \delta} + O(1) \right).$$

By Markov inequality $\Pr (|\mathbf{m}_k - E\mathbf{m}_k| \geq \epsilon) \leq Var (\mathbf{m}_k) / \epsilon^2$. Substituting $\epsilon = n^{-1} (p/n)^{ek/2(2+e) \delta} + O(1)$ we get

$$\Pr \left( |\mathbf{m}_k - E\mathbf{m}_k| \geq n^{-1} (p/n)^{ek/2(2+e) \delta} + O(1) \right) = O \left( n (n/p)^{ek/2(2+e) \delta} \right),$$

where in the last equality we used the estimate $Var (\mathbf{m}_k) = O \left( n^{-1} \right)$ (see Theorem [1.3]).

This concludes the proof of the theorem.

1.7. Proof of Theorem [1.4] Theorem [1.4] follows from Theorem [1.5] using the Markov inequality. In order to show that $\lim_{p \to \infty} \rho_E = \rho_{SC}$, it is enough to show that for every $k \in \mathbb{N}$ and $\delta > 0$ we have

$$\lim_{p \to \infty} \Pr (|\mathbf{m}_k - m_{SC,k}| \geq \delta) = 0.$$

The proof of the last assertion proceeds as follows: By the triangle inequality we have that $|\mathbf{m}_k - m_{SC,k}| \leq |\mathbf{m}_k - E\mathbf{m}_k| + |E\mathbf{m}_k - m_{SC,k}|$, therefore

$$\Pr (|\mathbf{m}_k - m_{SC,k}| \geq \delta) \leq \Pr (|\mathbf{m}_k - E\mathbf{m}_k| + |E\mathbf{m}_k - m_{SC,k}| \geq \delta).$$

By [1.3] there exists $P_0 \in \mathbb{N}$ such that $|E\mathbf{m}_k - m_{SC,k}| \leq \delta/2$, for every $p \geq P_0$, hence

$$\Pr (|\mathbf{m}_k - E\mathbf{m}_k| + |E\mathbf{m}_k - m_{SC,k}| \geq \delta) \leq \Pr (|\mathbf{m}_k - E\mathbf{m}_k| \geq \delta/2),$$

for every $p \geq P_0$. Now, using the Markov inequality

$$\Pr (|\mathbf{m}_k - E\mathbf{m}_k| \geq \delta/2) \leq \frac{Var (\mathbf{m}_k)}{\delta/2}.$$ 

This implies that

$$\Pr (|\mathbf{m}_k - \mathbf{m}_k^c| \geq \delta) \leq \frac{Var (\mathbf{m}_k)}{\delta/2} \xrightarrow{p \to \infty} 0,$$

where we use the estimate $Var (\mathbf{m}_k) = O \left( 1/n \right)$ (Equation [1.4]).

This concludes the proof of the theorem.

2. Proof of Theorem [1.6]

2.1. Preliminaries on matrix multiplication.
2.1.1. Paths.

Definition 2.1. A path of length \( k \) on a set \( A \) is a function \( \gamma : [0, k] \rightarrow A \). The path \( \gamma \) is called closed if \( \gamma(0) = \gamma(k) \). The path \( \gamma \) is called strict if \( \gamma(j) \neq \gamma(j + 1) \) for every \( j = 0, \ldots, k - 1 \).

Given a path \( \gamma : [0, k] \rightarrow A \), an element \( \gamma(j) \in A \) is called a vertex of the path \( \gamma \). A pair of consecutive vertices \( (\gamma(j), \gamma(j + 1)) \) is called an edge of the path \( \gamma \).

Let \( \mathcal{P}_k(A) \) denote the set of strict closed paths of length \( k \) on the set \( A \) and by \( \mathcal{P}_k(A, a, b) \) where \( a, b \in A \), the set of strict paths of length \( k \) on \( A \) which begin at the vertex \( a \) and end at the vertex \( b \).

Conventions:
- We will consider only strict paths and refer to these simply as paths.
- When considering a closed path \( \gamma \in \mathcal{P}_k(A) \), it will be sometime convenient to think of it as a function \( \gamma : \mathbb{Z}/k\mathbb{Z} \rightarrow A \).

2.1.2. Graphs associated with paths. Given a path \( \gamma \), we can associate to it an undirected graph \( G_\gamma = (V_\gamma, E_\gamma) \) where the set of vertices \( V_\gamma = \text{Im} \gamma \) and the set of edges \( E_\gamma \) consists of all sets \( \{a, b\} \subset A \) so that either \((a, b)\) or \((b, a)\) is an edge of \( \gamma \).

Remark 2.2. Since the graph \( G_\gamma \) is obtained from a path it is connected and, moreover, \( |V_\gamma|, |E_\gamma| \leq k \) where \( k \) is the length of \( \gamma \).

Definition 2.3. A closed path \( \gamma \in \mathcal{P}_k(A) \) is called a tree if the associated graph \( G_\gamma \) is a tree and every edge \( \{a, b\} \in E_\gamma \) is crossed exactly twice by \( \gamma \), once as \((a, b)\) and once as \((b, a)\).

Let \( T_k(A) \subset \mathcal{P}_k(A) \) denote the set of trees of length \( k \).

Remark 2.4. If \( \gamma \) is a tree of length \( k \) then \( k \) must be even, moreover, \( k = 2(|V_\gamma| - 1) \).

2.1.3. Isomorphism classes of paths. Let us denote by \( \Sigma(A) \) the permutation group \( \text{Aut}(A) \). The group \( \Sigma(A) \) acts on all sets which can be derived functorially from the set \( A \), in particular it acts on the set of closed paths \( \mathcal{P}_k(A) \) as follows: Given \( \sigma \in \Sigma(A) \) it sends a path \( \gamma : [0, k] \rightarrow A \) to \( \sigma \circ \gamma \).

An isomorphism class \( \tau = [\gamma] \in \mathcal{P}_k(A)/\Sigma(A) \) can be uniquely specified by a \( k + 1 \) ordered tuple of positive integers \((\tau_0, \ldots, \tau_k)\) where for each \( j \) the vertex \( \gamma(j) \) is the \( \tau_j \)th distinct vertex crossed by \( \gamma \). For example, the isomorphism class of the path \( \gamma = (a, b, c, a, b, a) \) is specified by \([\gamma] = (1, 2, 3, 1, 2, 1)\).

As a consequence we get that

\[
|\gamma| = |A|_{|V_\gamma|} = |A|(|A| - 1) \cdots (|A| - |V_\gamma| + 1).
\]

2.1.4. The combinatorics of matrix multiplication. First let us fix some general notations: If the set \( A \) is \([1, n]\) then we will denote
- \( \mathcal{P}_k = \mathcal{P}_k([1, n]) \), \( \mathcal{P}_k(i, j) = \mathcal{P}_k([1, n], i, j) \).
- \( T_k = T_k([1, n]) \).
- \( \Sigma_n = \Sigma([1, n]) \).

Let \( M \in \text{Mat}_{n \times n}(\mathbb{C}) \) be a matrix such that \( m_{ii} = 0 \), for every \( i \in [1, n] \). The \((i, j)\) entry \( m_{ij}^k \) of the \( k \)th power matrix \( M^k \) can be described as a sum of contributions indexed by strict paths, that is

\[
m_{ij}^k = \sum_{\gamma \in \mathcal{P}_k(i, j)} w_\gamma,
\]
where \( w_\gamma = m_{\gamma(0),\gamma(1)} \cdot m_{\gamma(1),\gamma(2)} \cdot \ldots \cdot m_{\gamma(k-1),\gamma(k)} \). Consequently, we can describe the trace of \( M^k \) as

\[
\text{Tr} (M^k) = \sum_{i \in [1,n]} \sum_{\gamma \in P_k(i,i)} w_\gamma = \sum_{\gamma \in P_k} w_\gamma,
\]

2.2. Fundamental estimates. Our goal here is to formulate the fundamental estimates that we will require for the proof of theorem 1.5.

Recall

\[
m_k = n^{-1} \text{Tr} (E^k) = n^{-1} \left( \frac{p}{n} \right)^{k/2} \text{Tr} \left( \left( G - I_n \right)^k \right).
\]

Since \((G - I_n)_{ii} = 0\) for every \( i \in [1,n] \) we can write, using Equation (2.2), the moment \( m_k \) in the form

\[
m_k = n^{-1} \left( \frac{p}{n} \right)^{k/2} \sum_{\gamma \in P_k} w_\gamma,
\]

where \( w_\gamma : \Omega_n \rightarrow \mathbb{C} \) is the random variable given by

\[
w_\gamma(S) = \langle S \circ \gamma(0), S \circ \gamma(1) \rangle \cdot \ldots \cdot \langle S \circ \gamma(k-1), S \circ \gamma(k) \rangle.
\]

Consequently, we get that

\[
E m_k = n^{-1} \left( \frac{p}{n} \right)^{k/2} \sum_{\gamma \in P_k} E w_\gamma.
\]

Lemma 2.5. Let \( \sigma \in \Sigma_n \) then \( E w_\gamma = E w_\sigma(\gamma) \). For a proof, see Appendix A.

Lemma 2.6 implies that the expectation \( E w_\gamma \) depends only on the isomorphism class \([\gamma]\) therefore we can write the sum (2.4) in the form

\[
E m_k = \sum_{\tau \in P_k/\Sigma_n} n^{-1} \left( \frac{p}{n} \right)^{k/2} |\tau| E w_\tau,
\]

where \( E w_\tau \) denotes the expectation \( E w_\gamma \) for any \( \gamma \in \tau \). Let us denote

\[
n(\tau) = n^{-1} \left( \frac{p}{n} \right)^{k/2} |\tau| = n^{-1} \left( \frac{p}{n} \right)^{k/2} n^{|V_\tau|} = p^{k/2} n^{|V_\tau| - 1 - k/2},
\]

where in the second equality we used (2.1). We conclude the previous development with the following formula

\[
E m_k = \sum_{\tau \in P_k/\Sigma_n} n(\tau) E w_\tau.
\]

Theorem 2.6 (Fundamental estimates). Let \( \tau \in P_k/\Sigma_n \).

(1) If \( k > 2 (|V_\tau| - 1) \) then

\[
\lim_{p \to \infty} n(\tau) E w_\tau = 0.
\]

(2) If \( k \leq 2 (|V_\tau| - 1) \) and \( \tau \) is not a tree then

\[
\lim_{p \to \infty} n(\tau) E w_\tau = 0.
\]

(3) If \( k \leq 2 (|V_\tau| - 1) \) and \( \tau \) is a tree then

\[
\lim_{p \to \infty} n(\tau) E w_\tau = 1.
\]

For a proof, see Subsection 2.4.
2.3. Proof of Theorem 1.5. The proof is a direct consequence of the fundamental estimates (Theorem 2.6).

2.3.1. Proof of Equation (1.3). Our goal is to show that \( \lim_{p \to \infty} \mathbb{E} m_k = m_{SC,k} \).

Using Equation (2.5) we can write

\[
(2.9) \quad \lim_{p \to \infty} \mathbb{E} m_k = \sum_{\tau \in \mathcal{P}_k/\Sigma_n} \lim_{p \to \infty} n(\tau) \mathbb{E} \mathbb{w}_\tau.
\]

When \( k \) is odd, no class \( \tau \in \mathcal{P}_k/\Sigma_n \) is a tree (see Remark 2.4), therefore by Theorem 2.6 all the terms in the right side of (2.9) are equal to zero, which implies that in this case \( \lim_{p \to \infty} \mathbb{E} m_k = 0 \). When \( k \) is even then, again by Theorem 2.6, only terms associated to trees yields a non-zero contribution to the right side of (2.9), therefore in this case

\[
\lim_{p \to \infty} \mathbb{E} m_k = \sum_{\tau \in \mathcal{T}_k/\Sigma_n} \lim_{p \to \infty} n(\tau) \mathbb{E} \mathbb{w}_\tau = \sum_{\tau \in \mathcal{T}_k/\Sigma_n} 1 = |\mathcal{T}_k|.
\]

For every \( m \in \mathbb{N} \), let \( \kappa_m \) denote the \( m \)th Catalan number, that is

\[
\kappa_m = \binom{2m}{m} \frac{1}{m+1}.
\]

On the one hand, the number of isomorphism classes of trees in \( \mathcal{T}_k/\Sigma_n \) can be described in terms of the Catalan numbers:

**Lemma 2.7.** If \( k = 2m \), \( m \in \mathbb{N} \) then

\[
|\mathcal{T}_{2m}| = \kappa_m.
\]

For a proof, see Appendix A.

On the other hand, the moments \( m_{SC,k} \) of the semicircle distribution are well-known and can be described in terms of the Catalan numbers as well:

**Lemma 2.8.** If \( k = 2m \) then \( m_{SC,k} = \kappa_m \); otherwise, if \( k \) is odd then \( m_{SC,k} = 0 \).

Consequently we obtain that for every \( k \in \mathbb{N} \)

\[
\lim_{p \to \infty} \mathbb{E} m_k = m_{SC,k}.
\]

This concludes the proof of the first part of the theorem.

2.3.2. Proof of Equation (1.4). By definition, \( \text{Var}(m_k) = \mathbb{E} m_k^2 - (\mathbb{E} m_k)^2 \).

Equation (2.3) implies that

\[
\mathbb{E} m_k^2 = n^{-2} \left( \frac{p}{n} \right)^k \sum_{\gamma_1, \gamma_2 \in \mathcal{P}_k} \mathbb{E} \left( \mathbb{w}_{\gamma_1} \mathbb{w}_{\gamma_2} \right),
\]

Equation (2.4) implies that

\[
(\mathbb{E} m_k)^2 = n^{-2} \left( \frac{p}{n} \right)^k \sum_{\gamma_1, \gamma_2 \in \mathcal{P}_k} \mathbb{E} w_{\gamma_1} \mathbb{E} w_{\gamma_2}
\]

When \( V_\gamma \cap V_{\gamma'} = \emptyset \), \( E(\mathbb{w}_{\gamma} \mathbb{w}_{\gamma'}) = E\mathbb{w}_{\gamma_1} E\mathbb{w}_{\gamma_2} \). If we denote by \( \mathcal{I}_k \subset \mathcal{P}_k \times \mathcal{P}_k \) the set of pairs \( (\gamma_1, \gamma_2) \) such that \( V_{\gamma_1} \cap V_{\gamma_2} \neq \emptyset \) then we can write

\[
\text{Var}(m_k) = n^{-2} \left( \frac{p}{n} \right)^k \sum_{(\gamma_1, \gamma_2) \in \mathcal{I}_k} \left( E(\mathbb{w}_{\gamma_1} \mathbb{w}_{\gamma_2}) - E\mathbb{w}_{\gamma_1} E\mathbb{w}_{\gamma_2} \right).
\]

The estimate of the variance now follows from
Lemma 2.9.

\[ n^{-2} (p/n)^k \sum_{(\gamma_1, \gamma_2) \in \mathcal{I}_k} |E(w_{\gamma_1} w_{\gamma_2})| = O(n^{-1}), \]

\[ n^{-2} (p/n)^k \sum_{(\gamma_1, \gamma_2) \in \mathcal{I}_k} |Ew_{\gamma_1}| |Ew_{\gamma_2}| = O(n^{-1}). \]

For a proof, see Appendix A.

This concludes the proof of the second part of the theorem.

2.4. Proof of Theorem 2.6. We begin by introducing notation: Given a set \( A \) we denote by \( \Omega (A) \) the set of injective maps

\[ \Omega (A) = \{ S : A \rightarrow \mathcal{D} \}, \]

and consider \( \Omega (A) \) as a probability space equipped with the uniform probability measure.

2.4.1. Proof of Equation (2.6). Let \( \tau = [\gamma] \in \mathcal{P}_k/\Sigma_n \) be an isomorphism class and assume that \( k > 2 (|V_\gamma| - 1). \) Our goal is to show that

\[ \lim_{p \rightarrow \infty} n (\tau) |Ew_\tau| = 0. \]

On the one hand, by Equation (2.10), we have that \(|[\gamma]| \sim n|V_\gamma|\), therefore

\[ n (\tau) = (p/n)^{k/2} |[\gamma]| \sim p^{k/2} n |V_\gamma| - 1 - k/2. \]

On the other hand

\[ Ew_\tau = |\Omega_n|^{-1} \sum_{S \in \Omega_n} w_\gamma (S) = |\Omega (V_\gamma)|^{-1} \sum_{S \in \Omega (V_\gamma)} w_\gamma (S). \]

By the triangle inequality, \(|Ew_\gamma| \leq |\Omega (V_\gamma)|^{-1} \sum_{S \in \Omega (V_\gamma)} |w_\gamma (S)|\), moreover, by the incoherence condition (Equation (1.1))

\[ |w_\gamma (S)| \leq \mu^{k/2} p^{-k/2}, \]

for every \( S \in \Omega (V_\gamma) \). In conclusion, we get that \(|Ew_\gamma| \leq \mu^{k/2} p^{-k/2}\) which combined with (2.10) yields

\[ n (\tau) |Ew_\tau| = O \left( n |V_\gamma|^{-1} - k/2 \right) p \rightarrow 0, \]

since, by assumption, \(|V_\gamma| - 1 - k/2 < 0. \)

This concludes the proof of Equation (2.6).

2.4.2. Proof of Equations (2.7) and (2.8). Let \( \tau = [\gamma] \in \mathcal{P}_k/\Sigma_n \) be an isomorphism class and assume that \( k \leq 2 (|V_\gamma| - 1). \) We prove Equations (2.7), (2.8) by induction on \(|V_\gamma|\).

Since \( k \leq 2 (|V_\gamma| - 1), \) there exists a vertex \( v = \gamma (i_0) \) where \( 0 \leq i_0 \leq k - 1, \) which is crossed once by the path \( \gamma. \) Let \( v_l = \gamma (i_0 - 1) \) and \( v_r = \gamma (i_0 + 1) \) be the adjacent vertices to \( v. \)

We will deal with the following two cases separately:

- Case 1. \( v_l \neq v_r, \)
- Case 2. \( v_l = v_r. \)
Introduce the following auxiliary constructions:

If \( v \neq v_r \), let \( \gamma_{\bar{v}} : [0, k - 1] \rightarrow [1, n] \) denote the closed path of length \( k - 1 \) defined by

\[
\gamma_{\bar{v}}(j) = \begin{cases} 
\gamma(j) & j \leq i_0 - 1 \\
\gamma(j + 1) & i_0 \leq j \leq k - 1 
\end{cases}.
\]

In words, the path \( \gamma_{\bar{v}} \) is obtained from \( \gamma \) by deleting the vertex \( v \) and inserting an edge connecting \( v_l \) to \( v_r \).

If \( v = v_r \), let \( \gamma_{\bar{v}} : [0, k - 2] \rightarrow [1, n] \) denote the closed path of length \( k - 2 \) defined by

\[
\gamma_{\bar{v}}(j) = \begin{cases} 
\gamma(j) & j \leq i_0 - 1 \\
\gamma(j + 2) & i_0 \leq j \leq k - 2 
\end{cases}.
\]

In words, the path \( \gamma_{\bar{v}} \) is obtained from \( \gamma \) by deleting the vertex \( v \) and identifying the vertices \( v_l \) and \( v_r \).

In addition, for every \( u \in V_{\gamma} - \{v, v_l, v_r\} \), let \( \gamma_u : [0, k] \rightarrow [1, n] \) denote the closed path of length \( k \) defined by

\[
\gamma_u(j) = \begin{cases} 
\gamma(j) & j \leq i_0 - 1 \\
u & j = i_0 \\
\gamma(j) & i_0 + 1 \leq j \leq k
\end{cases}.
\]

In words, the path \( \gamma_u \) is obtained from \( \gamma \) by deleting the vertex \( v \) and inserting an edge connecting \( v_l \) to \( u \) followed by an edge connecting \( u \) to \( v_r \).

**Important fact:** The number of vertices in the paths \( \gamma_{\bar{v}}, \gamma_u \) is \(|V_{\gamma}| - 1\).

The main technical statement is the following relation between the expectation \( Ew_{\gamma} \) and the expectations \( Ew_{\gamma_{\bar{v}}}, Ew_{\gamma_u} \).

**Proposition 2.10.**

\[
(2.11) \quad Ew_{\gamma} \sim p^{-1} Ew_{\gamma_{\bar{v}}} - (p |X|)^{-1} \sum_u Ew_{\gamma_u}.
\]

For a proof, see Appendix A

**Analysis of case 1.**

In this case the path \( \gamma \) is not a tree hence our goal is to show that

\[
\lim_{p \to \infty} n(\tau) Ew_{\tau} = 0.
\]

The length of \( \gamma_{\bar{v}} \) is \( k - 1 \) and \(|V_{\gamma_{\bar{v}}}| = |V_{\gamma}| - 1\), therefore

\[
n(\tau) \sim \frac{p^{k/2} n |V_{\gamma}|^{1-k/2}}{2 n^{1/2} n(|\gamma_{\bar{v}}|)}.
\]

The length of \( \gamma_u \) is \( k \) and \(|V_{\gamma_u}| = |V_{\gamma}| - 1\), therefore

\[
n(\tau) \sim \frac{p^{k/2} n |V_{\gamma}|^{1-k/2}}{n \cdot n(|\gamma_u|)}.
\]

Applying the above to (2.11) we obtain

\[
n(\tau) Ew_{\tau} \sim \frac{\left( n/p \right)^{1/2} n \left(|\gamma_{\bar{v}}|\right)}{\left(|\gamma_{\bar{v}}|\right)} Ew_{\gamma_{\bar{v}}}.
\]

By estimate (2.8) and the induction hypothesis \( n \left(|\gamma_{\bar{v}}|\right) Ew_{\gamma_{\bar{v}}} = O(1) \), therefore

\[
\lim_{p \to \infty} n(\tau) Ew_{\tau} = 0, \text{ since } (n/p) = o(1) \text{ (recall that we take } n = p^{1-\epsilon}).
\]

This concludes the proof of Equation (2.8).

**Analysis of case 2.**

The length of \( \gamma_{\bar{v}} \) is \( k - 2 \) and \(|V_{\gamma_{\bar{v}}}| = |V_{\gamma}| - 1\), therefore

\[
n(\tau) \sim \frac{p^{k/2} n |V_{\gamma}|^{1-k/2}}{n \cdot n \left(|\gamma_{\bar{v}}|\right)}.
\]
The length of $\gamma_u$ is $k$ and $|V_{\gamma_u}| = |V_{\gamma}| - 1$, therefore
$$n(\tau) \sim p^{k/2} n|V_{\gamma}|^{-1-k/2} \sim n(\gamma_u).$$

Applying the above to (2.11) yields
$$n(\tau) E w_{\tau} \sim n(\gamma_u) E w_{\gamma_u}.$$

If $\gamma$ is a tree then $\gamma_u$ is also a tree with a smaller number of vertices, therefore, by the induction hypothesis $\lim_{p \to \infty} n(\gamma_u) E w_{\gamma_u} = 1$ which implies by (2.11) that $\lim_{p \to \infty} n(\tau) E w_{\tau} = 1$ as well.

This concludes the proof of Equation (2.8).

3. Examples of incoherent dictionaries

3.1. Representation theory. We start with some preliminaries from representation theory of the finite Heisenberg group and the associated Weil representation (see [17] for a more detailed introduction).

3.1.1. The Heisenberg group. Let $(V, \omega)$ be a two-dimensional symplectic vector space over the finite field $\mathbb{F}_p$. The reader should think of $V$ as $\mathbb{F}_p^2$ with the standard symplectic form
$$\omega((\tau, w), (\tau', w')) = \tau w' - \tau' w.$$

Considering $V$ as an Abelian group, it admits a non-trivial central extension called the Heisenberg group. Concretely, the group $H$ can be presented as the set $H = V \times \mathbb{F}_p$ with the multiplication given by
$$(v, z) \cdot (v', z') = (v + v', z + z' + \frac{1}{2} \omega(v, v')).$$

The center of $H$ is $Z = Z(H) = \{(0, z) : z \in \mathbb{F}_p\}$. The symplectic group $Sp = Sp(V, \omega)$, which in this case is just isomorphic to $SL_2(\mathbb{F}_p)$, acts by automorphism of $H$ through its tautological action on the $V$-coordinate, that is, a matrix
$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

sends an element $(v, z)$, where $v = (\tau, w)$ to the element $(gv, z)$ where $gv = (a\tau + bw, c\tau + dw)$.

3.1.2. The Heisenberg representation. One of the most important attributes of the group $H$ is that it admits, principally, a unique irreducible representation. The precise statement goes as follows: Let $\psi: Z \to S^1$ be a non-degenerate unitary character of the center, for example, in this paper we take $\psi(z) = e^{2\pi i z}$. It is not difficult to show [27] that

**Theorem 3.1** (Stone-von Neuman). There exists a unique (up to isomorphism) irreducible unitary representation $\pi: H \to U(H)$ with central character $\psi$, that is, $\pi(z) = \psi(z) \cdot Id_H$, for every $z \in Z$.

The representation $\pi$ which appears in the above theorem will be called the Heisenberg representation.

The representation $\pi: H \to U(H)$ can be realized as follows: $H$ is the Hilbert space $\mathbb{C}(\mathbb{F}_p)$ of complex valued functions on the finite line, with the standard inner product
$$\langle f, g \rangle = \sum_{t \in \mathbb{F}_p} f(t) g(t),$$
for every \( f, g \in \mathbb{C}(\mathbb{F}_p) \), and the action \( \pi \) is given by

- \( \pi(\tau, 0)[f](t) = f(t + \tau) \);
- \( \pi(0, w)[f](x) = \psi(wt) f(t) \);
- \( \pi(z)[f](t) = \psi(z) f(t) \), \( z \in \mathbb{Z} \).

Here we are using \( \tau \) to indicate the first coordinate and \( w \) to indicate the second coordinate of \( V \simeq \mathbb{F}_p \times \mathbb{F}_p \).

We will call this explicit realization the standard realization.

### 3.1.3. The Weil representation

A direct consequence of Theorem 3.1 is the existence of a projective unitary representation \( \tilde{\rho} : Sp \to PU(H) \). The construction of \( \tilde{\rho} \) out of the Heisenberg representation \( \pi \) is due to Weil and it goes as follows:

Considering the Heisenberg representation \( \pi : H \to U(H) \) and an element \( g \in Sp \), one can define a new representation \( \pi^g : H \to U(H) \) by \( \pi^g(h) = \pi(g(h)) \). Clearly both \( \pi \) and \( \pi^g \) have the same central character \( \psi \) hence by Theorem 3.1 they are isomorphic. Since the space of intertwining morphisms \( \text{Hom}_H(\pi, \pi^g) \) is one dimensional, choosing for every \( g \in Sp \) a non-zero representative \( \tilde{\rho}(g) \in \text{Hom}_H(\pi, \pi^g) \) gives the required projective representation.

In more concrete terms, the projective representation \( \tilde{\rho} \) is characterized by the Egorov’s condition:

\[
\tilde{\rho}(g) \pi(h) \tilde{\rho}(g^{-1}) = \pi(g(h)),
\]

for every \( g \in Sp \) and \( h \in H \).

The important and non-trivial statement is that the projective representation \( \tilde{\rho} \) can be linearized in a unique manner into an honest unitary representation:

**Theorem 3.2.** There exists a unique \(^1\) unitary representation

\[
\rho : Sp \longrightarrow U(H),
\]

such that every operator \( \rho(g) \) satisfies Equation (3.1).

For the sake of concreteness, let us give an explicit description (which can be directly verified using Equation (3.1)) of the operators \( \rho(g) \), for different elements \( g \in Sp \), as they appear in the standard realization. The operators will be specified up to a unitary scalar.

- The standard diagonal subgroup \( A \subset Sp \) acts by (normalized) scaling: An element

  \[
a = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix},
\]

acts by

\[
S_a \ [f] \ (t) = \sigma(a) f(a^{-1}t),
\]

where \( \sigma : \mathbb{F}_p^\times \to \{\pm 1\} \) is the unique non-trivial quadratic character of the multiplicative group \( \mathbb{F}_p^\times \) (also called the Legendre character), given by \( \sigma(a) = a^{p-1}/2 \) (mod \( p \)).

- The subgroup of strictly lower diagonal elements \( U \subset Sp \) acts by quadratic exponents (chirps): An element

  \[
u = \begin{pmatrix} 1 & 0 \\ u & 1 \end{pmatrix},
\]

\(^1\)Unique, except in the case the finite field is \( \mathbb{F}_3 \).
acts by
\[ M_u [f] (t) = \psi(-\frac{ut^2}{2})f (t) . \]

• The Weyl element
\[ w = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \]
acts by discrete Fourier transform
\[ F [f] (w) = \frac{1}{\sqrt{p}} \sum_{t \in \mathbb{F}_p} \psi(wt)f (t) . \]

3.2. The Heisenberg dictionary. The Heisenberg dictionary is a collection of \( p + 1 \) orthonormal bases, each characterized, roughly, as eigenvectors of a specific linear operator. An elegant way to define this dictionary is using the Heisenberg representation \([19, 20]\).

3.2.1. Bases associated with lines. The Heisenberg group is non-commutative, yet it consists of various commutative subgroups which can be easily described as follows: Let \( L \subset V \) be a line in \( V \). One can associate to \( L \) a commutative subgroup \( A_L \subset H \), given by \( A_L = \{(l, 0) : l \in L \} \). It will be convenient to identify the group \( A_L \) with the line \( L \). Restricting the Heisenberg representation \( \pi \) to the commutative subgroup \( L \), namely, considering the restricted representation \( \pi : L \rightarrow U (H) \), one obtains a collection of operators \( \{ \pi(l) : l \in L \} \) which commute pairwisely. This, in turn, yields an orthogonal decomposition into character spaces
\[ \mathcal{H} = \bigoplus_{\chi} \mathcal{H}_\chi, \]
where \( \chi \) runs in the set \( \hat{L} \) of unitary characters of \( L \).

A more concrete way to specify the above decomposition is by choosing a non-zero vector \( l_0 \in L \). After such a choice, the character space \( \mathcal{H}_\chi \) naturally corresponds to the eigenspace of the linear operator \( \pi(l_0) \) associated with the eigenvalue \( \lambda = \chi (l_0) \).

It is not difficult to verify in this case that

**Lemma 3.3.** For every \( \chi \in \hat{L} \) we have \( \dim \mathcal{H}_\chi = 1 \).

Choosing a vector \( \varphi_\chi \in \mathcal{H}_\chi \) of unit norm \( \| \varphi_\chi \| = 1 \), for every \( \chi \in \hat{L} \) which appears in the decomposition, we obtain an orthonormal basis which we denote by \( B_L \).

**Theorem 3.4 (19, 20).** For every pair of different lines \( L, M \subset V \) and for every \( \varphi \in B_L, \phi \in B_M \)
\[ |\langle \varphi, \phi \rangle| = \frac{1}{\sqrt{p}} . \]

Since there exist \( p + 1 \) different lines in \( V \), we obtain in this manner a collection of \( p + 1 \) orthonormal bases
\[ \mathcal{D}_H = \coprod_{L \subset V} B_L. \]
which are \( \mu = 1 \)-coherent. We will call this dictionary, for obvious reasons, the Heisenberg dictionary.
3.3. The oscillator dictionary. Reflecting back on the Heisenberg dictionary we see that it consists of a collection of orthonormal bases characterized in terms of commutative families of unitary operators where each such family is associated with a commutative subgroup in the Heisenberg group $H$, via the Heisenberg representation $\pi : H \to U(\mathcal{H})$. In comparison, the oscillator dictionary \cite{16, 18} is characterized in terms of commutative families of unitary operators which are associated with commutative subgroups in the symplectic group $Sp$ via the Weil representation $\rho : Sp \to U(\mathcal{H})$.

3.3.1. Maximal tori. The commutative subgroups in $Sp$ that we consider are called maximal algebraic tori \cite{3} (not to be confused with the notion of a topological torus). A maximal (algebraic) torus in $Sp$ is a maximal commutative subgroup which becomes diagonalizable over some field extension. The most standard example of a maximal algebraic torus is the standard diagonal torus

$$A = \left\{ \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} : a \in \mathbb{F}_p^\times \right\}.$$ 

Standard linear algebra shows that up to conjugation\footnote{Two elements $h_1, h_2$ in a group $G$ are called conjugated elements if there exists an element $g \in G$ such that $g \cdot h_1 \cdot g^{-1} = h_2$. More generally, two subgroups $H_1, H_2 \subseteq G$ are called conjugated subgroups if there exists an element $g \in G$ such that $g \cdot H_1 \cdot g^{-1} = H_2$.} there exist two classes of maximal (algebraic) tori in $Sp$. The first class consists of those tori which are diagonalizable already over $\mathbb{F}_p$, namely, those are tori $T$ which are conjugated to the standard diagonal torus $A$ or more precisely such that there exists an element $g \in Sp$ so that $g \cdot T \cdot g^{-1} = A$. A torus in this class is called a split torus.

The second class consists of those tori which become diagonalizable over the quadratic extension $\mathbb{F}_{p^2}$, namely, those are tori which are not conjugated to the standard diagonal torus $A$. A torus in this class is called a non-split torus (sometimes it is called inert torus).

All split (non-split) tori are conjugated to one another, therefore the number of split tori is the number of elements in the coset space $Sp/N$ (see \cite{1} for basics of group theory), where $N$ is the normalizer group of $A$; we have

$$\#(Sp/N) = \frac{p(p + 1)}{2},$$

and the number of non-split tori is the number of elements in the coset space $Sp/M$, where $M$ is the normalizer group of some non-split torus; we have

$$\#(Sp/M) = p(p - 1).$$

Example of a non-split maximal torus. It might be suggestive to explain further the notion of non-split torus by exploring, first, the analogue notion in the more familiar setting of the field $\mathbb{R}$. Here, the standard example of a maximal non-split torus is the circle group $SO(2) \subseteq SL_2(\mathbb{R})$. Indeed, it is a maximal commutative subgroup which becomes diagonalizable when considered over the extension field $\mathbb{C}$ of complex numbers. The above analogy suggests a way to construct examples of maximal non-split tori in the finite field setting as well.

Let us assume for simplicity that $-1$ does not admit a square root in $\mathbb{F}_p$ or equivalently that $p \equiv 1 \text{ mod } 4$. The group $Sp$ acts naturally on the plane $V = \mathbb{F}_p^2$.
Consider the standard symmetric form $B$ on $V$ given by

$$B((x, y), (x', y')) = xx' + yy'.$$

An example of maximal non-split torus is the subgroup $SO = SO(V, B) \subset Sp$ consisting of all elements $g \in Sp$ preserving the form $B$, namely $g \in SO$ if and only if $B(gu, gv) = B(u, v)$ for every $u, v \in V$. In coordinates, $SO$ consists of all matrices $A \in SL_2(F_p)$ which satisfy $AA^t = I$. The reader might think of $SO$ as the "finite circle".

### 3.3.2. Bases associated with maximal tori.

Restricting the Weil representation to a maximal torus $T \subset Sp$ yields an orthogonal decomposition into character spaces

$$\mathcal{H} = \bigoplus_\chi \mathcal{H}_\chi,$$

where $\chi$ runs in the set $\hat{T}$ of unitary characters of the torus $T$.

A more concrete way to specify the above decomposition is by choosing a generator $t_0 \in T$, that is, an element such that every $t \in T$ can be written in the form $t = t_0^n$, for some $n \in \mathbb{N}$. After such a choice, the character spaces $\mathcal{H}_\chi$ which appears in (3.3) naturally corresponds to the eigenspace of the linear operator $\rho(t_0)$ associated to the eigenvalue $\lambda = \chi(t_0)$.

The decomposition (3.3) depends on the type of $T$ in the following manner (for details see [15]):

- In the case where $T$ is a split torus we have $\dim \mathcal{H}_\chi = 1$ unless $\chi = \sigma$, where $\sigma : T \to \{\pm 1\}$ is the unique non-trivial quadratic character of $T$ (also called the Legendre character of $T$), in the latter case $\dim \mathcal{H}_\sigma = 2$.
- In the case where $T$ is a non-split torus then $\dim \mathcal{H}_\chi = 1$ for every character $\chi$ which appears in the decomposition, in this case the quadratic character $\sigma$ does not appear in the decomposition.

Choosing for every character $\chi \in \hat{T}, \chi \neq \sigma$, a vector $\varphi_\chi \in \mathcal{H}_\chi$ of unit norm, we obtain an orthonormal system of vectors $B_T = \{\varphi_\chi : \chi \neq \sigma\}$.

**Important fact:** In the case when $T$ is a non-split torus, the set $B_T$ is an orthonormal basis.

**Example 3.5.** It would be beneficial to describe explicitly the system $B_A$ when $A \simeq G_m$ is the standard diagonal torus. The torus $A$ acts on the Hilbert space $\mathcal{H}$ by scaling (see Equation (3.2)).

For every $\chi \neq \sigma$, define a function $\varphi_\chi \in \mathbb{C}(F_p)$ as follows:

$$\varphi_\chi(t) = \begin{cases} \frac{1}{\sqrt{p}} \chi(t) & t \neq 0 \\ 0 & t = 0 \end{cases}.$$

It is easy to verify that $\varphi_\chi$ is a character vector with respect to the action $\rho : A \to U(\mathcal{H})$ associated to the character $\chi \cdot \sigma$. Concluding, the orthonormal system $B_A$ is the set $\{\varphi_\chi : \chi \in \hat{G}_m, \chi \neq \sigma\}$.

**Theorem 3.6 (16).** Let $\phi \in B_{T_1}$ and $\varphi \in B_{T_2}$

$$|\langle \phi, \varphi \rangle| \leq \frac{4}{\sqrt{p}}.$$
Since there exist \( p(p - 1) \) distinct non-split tori in \( Sp \), we obtain in this manner a collection of \( p(p - 1) \) orthonormal bases

\[
\mathcal{D}_O = \bigsqcup_{T \subset Sp} \text{non-split} B_T.
\]

which are \( \mu = 4 \)-coherent. We will call this dictionary the Oscillator dictionary.

### 3.4. The extended oscillator dictionary.

#### 3.4.1. The Jacobi group.

Let us denote by \( J \) the semi-direct product of groups

\[
J = Sp \rtimes H.
\]

The group \( J \) will be referred to as the Jacobi group.

#### 3.4.2. The Heisenberg-Weil representation.

The Heisenberg representation \( \pi : H \rightarrow U(H) \) and the Weil representation \( \rho : Sp \rightarrow U(H) \) combine to a representation of the Jacobi group

\[
\tau = \rho \rtimes \pi : J \rightarrow U(H),
\]

defined by \( \tau(g, h) = \rho(g) \pi(h) \). The fact that \( \tau \) is indeed a representation is a direct consequence of the Egorov’s condition - Equation (3.1). We will refer to the representation \( \tau \) as the Heisenberg-Weil representation.

#### 3.4.3. Maximal tori in the Jacobi group.

Given a non-split torus \( T \subset Sp \), the conjugate subgroup \( T_v = vTv^{-1} \subset J \) for every \( v \in V \) (the multiplication is in the group \( J \)), will be called a maximal non-split torus in \( J \).

It is easy to verify that the subgroups \( T_v, T_u \) are distinct for \( v \neq u \); moreover, for different tori \( T \neq T' \subset Sp \) the subgroups \( T_v, T'_u \) are distinct for every \( v, u \in V \). This implies that there are \( p(p - 1)p^2 \) non-split maximal tori in \( J \).

#### 3.4.4. Bases associated with maximal tori.

Restricting the Heisenberg-Weil representation \( \tau \) to a maximal non-split torus \( T \subset J \) yields a basis \( B_T \) consisting of character vectors. A way to think of the basis \( B_T \) is as follows: If \( T = T_v \) where \( T \) is a maximal torus in \( Sp \) then the basis \( B_T \) can be derived from the already known basis \( B_T \) by

\[
B_{T_v} = \pi(v) B_T,
\]

namely, the basis \( B_{T_v} \) consists of the vectors \( \pi(v) \varphi \) where \( \varphi \in B_T \).

Interestingly, given any two tori \( T_1, T_2 \subset J \), the bases \( B_{T_1}, B_{T_2} \) remain \( \mu = 4 \)-coherent - this is a direct consequence of the following generalization of Theorem 3.6.

**Theorem 3.7 (10).** Given (not necessarily distinct) tori \( T_1, T_2 \subset Sp \) and a pair of distinct vectors \( \varphi \in B_{T_1}, \phi \in B_{T_2} \)

\[
|\langle \varphi, \pi(v) \phi \rangle| \leq \frac{4}{\sqrt{p}},
\]

for every \( v \in V \).

Since there exist \( p(p - 1)p^2 \) distinct non-split tori in \( J \), we obtain in this manner a collection of \( p(p - 1)p^2 \) orthonormal bases

\[
\mathcal{D}_{EO} = \bigsqcup_{T \subset J} \text{non-split} B_T.
\]
which are $\mu = 4$-coherent. We will call this dictionary the *extended oscillator dictionary*.

**Remark 3.8.** A way to interpret Theorem 3.7 is to say that any two different vectors $\varphi \neq \phi \in \mathcal{D}_O$ are incoherent in a stable sense, that is, their coherency is $4/\sqrt{p}$ no matter if any one of them undergoes an arbitrary time/phase shift. This property seems to be important in communication where a transmitted signal may acquire time shift due to asynchronous communication and phase shift due to Doppler effect.

**Appendix A. Proof of statements**

**A.1. Proof of Lemma 2.5.** Let $\gamma \in \mathcal{P}_k$ and $\sigma \in \Sigma_n$. By definition, $\sigma (\gamma) = \sigma \circ \gamma : [0,k] \to [1,n]$. Write

$$Ew_{\sigma(\gamma)} = |\Omega_n|^{-1} \sum_{S \in \Omega_n} w_{\sigma(\gamma)} (S).$$

Direct verification reveals that $w_{\sigma(\gamma)} (S) = w_{\gamma} (\sigma (S))$ where $\sigma (S) = S \circ \sigma : [1,n] \to \mathcal{D}$, hence

$$\sum_{S \in \Omega_n} w_{\sigma(\gamma)} (S) = \sum_{S \in \Omega_n} w_{\gamma} (\sigma (S)) = \sum_{S \in \Omega_n} w_{\gamma} (S),$$

which implies that $Ew_{\sigma(\gamma)} = Ew_{\gamma}$.

This concludes the proof of the lemma.

**A.2. Proof of Lemma 2.7.** We need to introduce the notion of a *Dick word*.

**Definition A.1.** A Dick work of length $2m$ is a sequence $D = d_1d_2...d_{2m}$ where $d_i = \pm 1$, which satisfies

$$\sum_{i=1}^{l} d_i \geq 0,$

for every $l = 1,..,2m$.

Let us the denote by $\mathcal{D}_{2m}$ the set of Dick words of length $2m$. It is well know that $|\mathcal{D}_{2m}| = \kappa_m$.

In addition, let us denote by $\mathcal{T}_{2m} \subset \mathcal{P}_{2m}$ the subset of trees of length $2m$. Our goal is to establish a bijection

$$D : \mathcal{T}_{2m}/\Sigma_n \cong \mathcal{D}_{2m}.$$

Given a tree $\gamma \in \mathcal{T}_{2m}$ define the word $D (\gamma) = d_1d_2...d_{2m}$ as follows:

$$d_i = \begin{cases} 1 & \text{if } \gamma (i-1) \text{ is crossed for the first time on the } i-1 \text{ step} \\ -1 & \text{otherwise} \end{cases}$$

The word $D (\gamma)$ is a Dick word since $\sum_{i=1}^{l} d_i$ counts the number of vertices visited exactly once by $\gamma$ in the first $l$ steps, therefore, it is greater or equal to zero.

On the one direction, if two trees $\gamma_1, \gamma_2$ are isomorphic then $D (\gamma_1) = D (\gamma_2)$. In addition, it is easy to verify that the tree $\gamma$ can be reconstructed from the pair $(D (\gamma), \overline{V}_\gamma)$ where $\overline{V}_\gamma$ is the set of vertices of $\gamma$ equipped with the following linear order:

$$v < u \iff \gamma \text{ crosses } v \text{ for the first time before it crosses } u \text{ for the first time}.$$
This implies that $D$ defines an injection from $\mathcal{I}_{2m}/\Sigma_n$ into $D_{2m}$.

Conversely, it is easy to verify that for every Dick word $D \in D_{2m}$ there is a tree $\gamma \in \mathcal{P}_{2k}$ such that $D = D(\gamma)$, which implies that the map $D$ is surjective.

This concludes the proof of the lemma.

A.3. **Proof of Lemma 2.9.** We begin with an auxiliary construction. Define a map $\sqcup : \mathcal{I}_k \to \mathcal{P}_{2k}$ as follows: Given $(\gamma_1, \gamma_2) \in \mathcal{I}_k$, let $0 \leq i_1 \leq k$ be the first index so that $\gamma_1(i_1) \in V_{\gamma_2}$ and let $0 \leq i_2 \leq k$ be the first index such that $\gamma_1(i_1) = \gamma_2(i_2)$.

Define

$$
\gamma_1 \sqcup \gamma_2(j) = \begin{cases} 
\gamma_1(j) & 0 \leq j \leq i_1 \\
\gamma_2(i_2 - i_1 + j) & i_1 \leq j \leq i_1 + k \\
\gamma_1(j - k) & i_1 + k \leq j \leq 2k
\end{cases}
$$

In words, the path $\gamma_1 \sqcup \gamma_2$ is obtained, roughly, by substituting the path $\gamma_2$ instead of the vertex $\gamma_1(i_1)$. Clearly, the map $\sqcup$ is injective and commutes with the action of $\Sigma_n$, therefore, we get, in particular, that the number of elements in the isomorphism class $[\gamma_1, \gamma_2] \in \mathcal{I}_k/\Sigma_n$ is smaller or equal than the number of elements in the isomorphism class $[\gamma_1 \sqcup \gamma_2] \in \mathcal{P}_{2k}/\Sigma_n$

(A.1) 
$$
||[\gamma_1, \gamma_2]| | \leq \|[\gamma_1 \sqcup \gamma_2]\|.
$$

**First estimate.** We need to show

$$
n^{-2} \left(\frac{p}{n}\right)^k \sum_{(\gamma_1, \gamma_2) \in \mathcal{I}_k} |E(w_{\gamma_1}w_{\gamma_2})| = O\left(n^{-1}\right).
$$

Write

$$
\sum_{(\gamma_1, \gamma_2) \in \mathcal{I}_k} |E(w_{\gamma_1}w_{\gamma_2})| = \sum_{[\gamma_1, \gamma_2] \in \mathcal{I}_k/\Sigma_n} \|[\gamma_1, \gamma_2]| \cdot |E(w_{\gamma_1}w_{\gamma_2})|.
$$

It is enough to show that for every $[\gamma_1, \gamma_2] \in \mathcal{I}_k/\Sigma_n$

(A.2) 
$$
n^{-2} \left(\frac{p}{n}\right)^k \|[\gamma_1, \gamma_2]| \cdot |E(w_{\gamma_1}w_{\gamma_2})| = O\left(n^{-1}\right).
$$

Fix an isomorphism class $[\gamma_1, \gamma_2] \in \mathcal{I}_k/\Sigma_n$. By Equation (A.1) we have that $||[\gamma_1, \gamma_2]| | \leq \|[\gamma_1 \sqcup \gamma_2]\|$. In addition, a simple observation reveals that $w_{\gamma_1}w_{\gamma_2} = w_{\gamma_1 \sqcup \gamma_2}$ which implies that $E \left( w_{\gamma_1}, w_{\gamma_2} \right) = E \left( w_{\gamma_1 \sqcup \gamma_2} \right)$. In conclusion, since the length of $\gamma_1 \sqcup \gamma_2$ is $2k$, we get that

$$
n^{-2} \left(\frac{p}{n}\right)^k ||[\gamma_1, \gamma_2]| | \cdot |E(w_{\gamma_1}w_{\gamma_2})| \leq n^{-1}n \left([\gamma_1 \sqcup \gamma_2] \right) |E\left( w_{\gamma_1 \sqcup \gamma_2} \right)|.
$$

Finally, by Theorem 2.6 we have that $n \left([\gamma_1 \sqcup \gamma_2] \right) |E\left( w_{\gamma_1 \sqcup \gamma_2} \right)| = O\left(1\right)$, hence, Equation (A.2) follows.

This concludes the proof of the first estimate.

**Second estimate.** We need to show

$$
n^{-2} \left(\frac{p}{n}\right)^k \sum_{(\gamma_1, \gamma_2) \in \mathcal{I}_k} |Ew_{\gamma_1}| |Ew_{\gamma_2}| = O\left(n^{-1}\right).
$$

Write

$$
\sum_{(\gamma_1, \gamma_2) \in \mathcal{I}_k} |Ew_{\gamma_1}| |Ew_{\gamma_2}| = \sum_{[\gamma_1, \gamma_2] \in \mathcal{I}_k/\Sigma_n} \|[\gamma_1, \gamma_2]| \cdot |Ew_{\gamma_1}| |Ew_{\gamma_2}|.
$$

It is enough to show that for every $[\gamma_1, \gamma_2] \in \mathcal{I}_k/\Sigma_n$

(A.3) 
$$
n^{-2} \left(\frac{p}{n}\right)^k \|[\gamma_1, \gamma_2]| \cdot |Ew_{\gamma_1}| |Ew_{\gamma_2}| = O\left(n^{-1}\right).
$$
Fix an isomorphism class $[\gamma_1, \gamma_2] \in \mathcal{I}_k / \Sigma_n$. By Equation (A.1) we have that $||\gamma_1, \gamma_2|| \leq ||\gamma_1 \cup \gamma_2||$. For every path $\gamma$, we have that $||\gamma|| = n(|V_\gamma|) \sim n|V_{\gamma_1}|$ (since always $|V_\gamma| \leq k$ and we assume that $k$ is fixed, that is, it does not depend on $p$), in particular

$$
||\gamma_1|| \sim n|V_{\gamma_1}|,
$$

$$
||\gamma_2|| \sim n|V_{\gamma_2}|,
$$

$$
||\gamma_1 \cup \gamma_2|| \sim n|V_{\gamma_1 \cup \gamma_2}|.
$$

By construction, $|V_{\gamma_1 \cup \gamma_2}| \leq |V_{\gamma_1}| + |V_{\gamma_2}| - 1$ (we assume that $V_{\gamma_1} \cap V_{\gamma_2} \neq \emptyset$), therefore, $||\gamma_1 \cup \gamma_2|| = O\left(n^{-1} ||\gamma_1|| ||\gamma_2||\right)$. In conclusion, we get that

$$
n^{-2} \left(\frac{p}{n}\right)^k ||\gamma_1, \gamma_2|| \cdot |Ew_{\gamma_1}| |Ew_{\gamma_2}| = O\left(n^{-2} (n(\gamma_1) n(\gamma_2) |Ew_{\gamma_1}| |Ew_{\gamma_2}|\right),
$$

where we used the identity $n^{-2} \left(\frac{p}{n}\right)^k ||\gamma_1|| ||\gamma_2|| = n(\gamma_1) n(\gamma_2)$. Finally, by Theorem 2.4, we have that $n(\gamma_i) |Ew_{\gamma_i}| = O(1)$, $i = 1, 2$, hence, Equation (A.3) follows.

This concludes the proof of the second estimate and concludes the proof of the lemma.

A.4. Proof of Proposition 2.10

Write

$$
Ew_\gamma = |\Omega(V_\gamma)|^{-1} \sum_{S \in \Omega(V_\gamma)} w_\gamma(S)
$$

(A.4)

$$
= |\Omega(V_\gamma)|^{-1} \sum_{S \in \Omega(V_\gamma)} \sum_{b \in \mathcal{D} \setminus S(V_\gamma \setminus \{v\})} w_\gamma(S \cup b),
$$

where $S \cup b : V_\gamma \rightarrow \mathcal{D}$ is given by

$$
S \cup b(u) = \begin{cases} 
S(u) & u \neq v \\
b & u = v
\end{cases}.
$$

Write

$$
\sum_{b \in \mathcal{D} \setminus S(V_\gamma \setminus \{v\})} w_\gamma(S \cup b) = \sum_{b \in \mathcal{D}} w_\gamma(S \cup b) - \sum_{b \in S(V_\gamma \setminus \{v\})} w_\gamma(S \cup b).
$$

Let us analyze separately the two terms in right side of the above equation.

First term.

Write

$$
\sum_{b \in \mathcal{D}} w_\gamma(S \cup b) = \sum_{x \in \mathcal{X}} \sum_{b_x \in B_x} w_\gamma(S \cup b_x).
$$

Furthermore

$$
w_\gamma(S \cup b_x) = \langle , .., \rangle \langle S(v_1), b_x \rangle \langle b_x, S(v_r) \rangle .. \langle , .., \rangle,
$$

Since $B_x$ is an orthonormal basis

$$
\sum_{b_x \in B_x} \langle S(v_1), b_x \rangle \langle b_x, S(v_r) \rangle = \langle S(v_1), S(v_r) \rangle,
$$

which implies that $\sum_{b_x \in B_x} w_\gamma(S \cup b_x) = w_{\gamma_0}(S)$. Concluding, we obtain

(A.5)

$$
\sum_{b \in \mathcal{D}} w_\gamma(S \cup b) = |\mathcal{X}| w_{\gamma_0}(S).
$$
Second term.
Let $b \in S(V_\gamma \setminus \{v\})$. Since $S$ is injective, there exists a unique $u \in V_\gamma \setminus \{v\}$ such that $b = S(u)$, therefore

$$w_\gamma (S \sqcup b) = (\cdot, \cdot) \cdot (S(v_1), b) \cdot (S(v_r), b) \cdot (\cdot, \cdot) = w_{\gamma_u} (S).$$

Furthermore, observe that when $u = v_1$ or $u = v_r$ we have that $w_{\gamma_u} (S) = w_{\gamma_v} (S)$. In conclusion, we obtain

$$\sum_{b \in S(V_\gamma \setminus \{v\})} w_\gamma (S \sqcup b) = 2w_{\gamma_v} (S) + \sum_{u \in V_\gamma \setminus \{v_1, v_r, v\}} w_{\gamma_u} (S).$$

Combining (A.5) and (A.6) yields

$$\sum_{b \in \mathcal{D} \setminus S(V_\gamma \setminus \{v\})} w_\gamma (S \sqcup b) = (|\mathcal{X}| - 2) w_{\gamma_v} (S) - \sum_{u \in V_\gamma \setminus \{v_1, v_r, v\}} w_{\gamma_u} (S).$$

Substituting the above in (A.4) yields

$$E w_\gamma = (|\mathcal{X}| - 2)|\Omega(V_\gamma)|^{-1} \sum_{S \in \Omega(V_\gamma \setminus \{v\})} w_{\gamma_v} (S) - \sum_{u \in V_\gamma \setminus \{v_1, v_r, v\}} |\Omega(V_v)|^{-1} \sum_{S \in \Omega(V_\gamma \setminus \{v\})} w_{\gamma_u} (S).$$

Finally, direct counting argument reveals that

$$|\Omega(V_v)| \sim p |\mathcal{X}| |\Omega(V_v)|,$$

$$|\Omega(V_v)| \sim p |\mathcal{X}| |\Omega(V_u)|.$$}

Hence (A.7) yields

$$E w_\gamma \sim p^{-1} E w_{\gamma_v} - \sum_{u \in V_\gamma \setminus \{v_1, v_r, v\}} (p |\mathcal{X}|)^{-1} E w_{\gamma_u}.$$

This concludes the proof of the proposition.

References

[1] Artin M., Algebra. Prentice Hall, Inc., Englewood Cliffs, NJ (1991).
[2] Applebaum L., Howard S., Searle S., and Calderbank R., Chirp sensing codes: Deterministic compressed sensing measurements for fast recovery. (Preprint, 2008).
[3] Borel A. Linear algebraic groups. Graduate Texts in Mathematics, 126. Springer-Verlag, New York (1991).
[4] Baraniuk R., Davenport M., DeVore R.A. and Wakin M.B., A simple proof of the restricted isometry property for random matrices. Constructive Approximation, to appear (2007).
[5] Bruckstein A.M., Donoho D.L. and Elad M., “From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images”, to appear in SIAM Review (2007).
[6] Compressive Sensing Resources. Available at http://www dsp.ece.rice.edu/cs/.
[7] Candès E. Compressive sampling. In Proc. International Congress of Mathematicians, vol. 3, Madrid, Spain (2006).
[8] Candès E., Romberg J. and Tao T., Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. Information Theory, IEEE Transactions on, vol. 52, no. 2, pp. 489–509 (2006).
[9] Candès E., and Tao T., Decoding by linear programming. IEEE Trans. on Information Theory, 51(12), pp. 4203 - 4215 (2005).
[10] Donoho D., Compressed sensing. IEEE Transactions on Information Theory, vol. 52, no. 4, pp. 1289–1306 (2006).
[11] Donoho D.L. and Elad M., Optimally sparse representation in general (non-orthogonal) dictionaries via $\ell_1$ minimization. Proc. Natl. Acad. Sci. USA 100, no. 5, 2197–2202 (2003).
[12] DeVore R. A., Deterministic constructions of compressed sensing matrices. J. Complexity 23 (2007), no. 4-6, 918–925.
[13] Daubechies I., Grossmann A. and Meyer Y., Painless non-orthogonal expansions. J. Math. Phys., 27 (5), pp. 1271-1283 (1986).
[14] Elad M. and Bruckstein A.M., A Generalized Uncertainty Principle and Sparse Representation in Pairs of Bases. IEEE Trans. On Information Theory, Vol. 48, pp. 2558-2567 (2002).
[15] Gurevich S. and Hadani R., On the diagonalization of the discrete Fourier transform. Applied and Computational Harmonic Analysis. To appear (2008).
[16] Gurevich S., Hadani R. and Sochen N., The finite harmonic oscillator and its associated sequences. Proceedings of the National Academy of Sciences of the United States of America, in press (2008).
[17] Gurevich S., Hadani R., Sochen N., On some deterministic dictionaries supporting sparsity . Special issue on sparsity, the Journal of Fourier Analysis and Applications. To appear (2008).
[18] Gurevich S., Hadani R., Sochen N., The finite harmonic oscillator and its applications to sequences, communication and radar. IEEE Transactions on Information Theory, vol. 54, no. 9. September 2008.
[19] Howe R., Nice error bases, mutually unbiased bases, induced representations, the Heisenberg group and finite geometries. Indag. Math. (N.S.) 16 , no. 3-4, 553–583 (2005).
[20] Howard S. D., Calderbank A. R. and Moran W. The finite Heisenberg-Weyl groups in radar and communications. EURASIP J. Appl. Signal Process. (2006).
[21] Howard S.D., Calderbank A.R., and Searle S.J., A fast reconstruction algorithm for deterministic compressive sensing using second order Reed-Muller codes. CISS (2007).
[22] Indyk P., Explicit constructions for compressed sensing of sparse signals. SODA (2008).
[23] Jafarpour S., Efficient Compressed Sensing using Lossless Expander Graphs with Fast Bilateral Quantum Recovery Algorithm. arXiv:0806.3799 (2008).
[24] Jafarpour S., Xu W., Hassibi B., Calderbank R., Efficient and Robust Compressive Sensing using High-Quality Expander Graphs. Submitted to the IEEE transaction on Information Theory (2008).
[25] Xu W. and Hassibi B., Efficient Compressive Sensing with Deterministic Guarantees using Expander Graphs. Proceedings of IEEE Information Theory Workshop, Lake Tahoe (2007).
[26] Saligrama V., Deterministic Designs with Deterministic Guarantees: Toeplitz Compressed Sensing Matrices, Sequence Designs and System Identification. arXiv:0806.4958 (2008).
[27] Terras A., Fourier analysis on finite groups and applications. London Mathematical Society Student Texts, 43. Cambridge University Press, Cambridge (1999).
[28] Tropp J.A., On the conditioning of random subdictionaries. Appl. Comput. Harmonic Anal., vol. 25, pp. 1–24, 2008.
[29] Tropp J.A., Norms of random submatrices and sparse approximation. Submitted to Comptes- Rendus de l’Académie des Sciences (2008).
[30] Weil A., Sur certains groupes d’opérateurs unitaires. Acta Math. 111, 143-211 (1964).

Department of Mathematics, University of California, Berkeley, CA 94720, USA.  
E-mail address: shamgar@math.berkeley.edu

Department of Mathematics, University of Chicago, IL 60637, USA.  
E-mail address: hadani@math.uchicago.edu