NEAR-OPTIMAL CODED APERTURES FOR IMAGING VIA NAZAROV’S THEOREM

Ganesh Ajjanagadde*  Christos Thrampoulidis†  Adam Yedidia*  Gregory Wornell*

*Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology
†Department of Electrical and Computer Engineering, University of California, Santa Barbara

ABSTRACT
We characterize the fundamental limits of coded aperture imaging systems up to universal constants by drawing upon a theorem of Nazarov regarding Fourier transforms. Our work is performed under a simple propagation and sensor model that accounts for thermal and shot noise, scene correlation, and exposure time. Focusing on mean square error as a measure of linear reconstruction quality, we show that appropriate application of a theorem of Nazarov leads to essentially optimal coded apertures, up to a constant multiplicative factor in exposure time. Additionally, we develop a heuristically efficient algorithm to generate such patterns that explicitly takes into account scene correlations. This algorithm finds apertures that correspond to local optima of a certain potential on the hypercube, yet are guaranteed to be tight. Finally, for i.i.d. scenes, we show improvements upon prior work by using spectrally flat sequences with bias. The development focuses on 1D apertures for conceptual clarity; the natural generalizations to 2D are also discussed.

Index Terms— coded aperture cameras, computational photography, optical signal processing, Fourier analysis

1. INTRODUCTION

Certain modern imaging systems, especially those operating at high frequencies, use coded apertures. In these systems, a spatial mask that selectively blocks light from reaching the sensor is used as opposed to a traditional lens. The scene is then recovered by suitable post-processing. Perhaps the earliest and simplest instance of coded aperture imaging is the pinhole structure; see, e.g., [1] for a survey. The development of X-ray and gamma-ray astronomy gave rise to more sophisticated coded apertures [2,3] to get around the lack of lenses and mirrors in such settings. Both proposed using random blockage patterns with a specified mean transmittance as a method to increase the aperture size as compared to the classical pinhole while retaining its resolution benefits.

More modern developments include the usage of uniformly redundant arrays (URA) to improve upon random on-off patterns [4], anti-pinhole imaging [5], as well as the combining of mask and lens in order to, e.g., facilitate depth estimation [6], deblur out-of-focus elements in an image [7], enable motion deblurring [8], and/or recover 4D lightfields [9]. Even more recent work seeks to forgo lenses altogether to decrease costs and meet physical constraints [10,11]. Understanding coded apertures is also relevant in non-line-of-sight applications where masks naturally occur as scene occlusions [12,13].

In light of the increased importance of coded apertures, prior work [14] described a model under which they can be analyzed.

This model uses far-field geometric optics to model light propagation and a sensor model that includes thermal and shot noise components. Together with mutual information (MI) as a performance metric, [14] compared the classical random on-off apertures [2,3] of varying intensity to the “spectrally flat” patterns with transmissivity 1/2 (same as the URA of [4]). Among other things, the analysis showed that when shot noise dominates thermal noise, randomly generated masks with lower transmissivity than 1/2 offered greater performance compared to spectrally flat patterns of transmissivity 1/2.

This paper extends the work of [14] in multiple respects that may be broadly grouped into the following three main contributions.

First, we refine the model of [14] by incorporating exposure time. Here, we analyze linearly-constrained minimum mean square error (LMMSE) estimation as opposed to MI given its direct operational relevance, though we remark in advance that our conclusions carry over to the MI criterion used in [14]; see Sec. 2.

Second, we remark upon the existence and construction of spectrally flat sequences with transmissivities 1/8, 1/4 in addition to 1/2. This extends the range of parameters where we have a sharp characterization of optimal coded apertures in our framework, and gives a tight answer to the problem of optimal coded apertures for i.i.d. scenes; see Props. 2,3 for precise statements.

Third, we provide optimal (up to a universal constant) coded apertures, both in 1D as well as in 2D, applicable for any prior on the spectrum of the scene at hand. The sense of tightness of the optimality is given precisely in Prop. 4. This includes (but is not limited to) the naturally occuring power law $\gamma$ (same as the URA of [4]). Among other things, the analysis showed that when shot noise dominates thermal noise, randomly generated masks with lower transmissivity than 1/2 offered greater performance compared to spectrally flat patterns of transmissivity 1/2.

We first describe our model, and discuss how it differs from that in [14]. We use the standard Poisson model of classical optics for photon counting, and emphasize its dependence on the exposure time $t$. The analysis of MI under Poisson models is cumbersome, and even with mean square error (MSE) it is often unclear how to achieve optimal MSE in practice. As such, the standard estimation process is linear; indeed, the work of [2,3] used correlation decoders. In fact, both [2,3] give beautiful analog realizations of such a decoder. Accordingly, we emphasize LMMSE. We note that if one used a Gaussian model instead, LMMSE is the same as MSE, and MSE is in turn essentially equivalent to MI in the low SNR limit [18,19]. LMMSE

This work was supported, in part, by DARPA under Contract No. HR0011-16-C-0030, and by NSF under Grant No. CCF-1816209.
depends purely on first and second moments, so in our mathematical study we do not emphasize the specific Poisson statistics.

We use a 1D model, as in [14], to simplify the exposition of the concepts and results. We emphasize that all of the results of this paper generalize naturally to the analogous 2D model, whose discussion we defer to Sec. [4]

Let $f$ denote the intensities of the unknown 1D scene of length $n$ of expected total power $J$. Let $\mathbb{E}[f] = (J/n)1$, $\mathbf{Cov}[f] = \mathbf{Q}$. We assume $\mathbf{Q}$ is circular and diagonalized as $\mathbf{Q} = \mathbf{F}_n^* \mathbf{D}_F \mathbf{F}_n$, where $\mathbf{F}_n$ is the unitary discrete Fourier transform (DFT) matrix and $\mathbf{D}_F = \text{diag}(d)$. The measurements at the imaging plane are denoted by $y_j, j \in [n]$ and the $n \times n$ transfer matrix $A$ models the aperture. We assume its entries all satisfy $0 \leq A_{ji} \leq 1/n$ to model that the light cannot be redirected, and $\sum_j A_{ji} \leq 1$ to model local conservation of power. An ideal, perfectly focused, lens may be treated in this setup by $A = I$, as it redirects light perfectly.

We assume $A_{ji} = (1/n) a_{i-j \text{ (mod } n)}$ for $a \geq 0$, i.e. $A$ is circular. Let $p(a) = (1/n) \sum_i a_i$ be the transmissivity of the aperture. The noise component is denoted by $z$ and its statistics are given by $\mathbb{E}[z] = 0$, $\mathbf{Cov}[z] = ((W + Jp)/n) \mathbf{I}$, where $W, J$ correspond to thermal and shot noise respectively, and $t$ is the exposure time. With these, our measurement model is then given by $y_j = t \sum_i A_{ji} f_i + z_j$, which leads to the following expression for the LMMSE of estimating $f$ by $y$:

$$m(n, t, W, J, d, a) = \frac{1}{\sum_{i=0}^{n-1} \frac{a_i^2}{d_i} + \frac{1}{(W + Jp(a))}}. \quad (1)$$

Here, $\hat{a}$ is the DFT of $a$. In general, we assume $d_i = (1/n)d(i/n)$ are $n$ equally spaced samples from a nonnegative, bounded, continuous function $d(x)$ on $[0, 1]$ with symmetry $d(x) = d(1 - x)$ and normalized so that $d(0) = 0$. For example, i.i.d. scenes correspond to $d(x) = \Theta$. Note that we may rewrite Prop. [1] holds in greater generality. The above restriction on the form of $d$ simply ensures correct physical scaling (invariant with respect to $n$) of the variance of total scene intensity coming from an arbitrary direction.

It is instructive to compare an ideal lens to a mask with respect to (1), as a function of exposure time $t$. An ideal lens satisfies $A = I$, (i.e., $a = (n, n, \ldots, n)$). Thus $\hat{a} = (n, n, \ldots, n)$. Then from (1), it can be readily seen that for a $t$ growing with $n$ (say $t \sim \log(n)$), the LMMSE decays to 0 as $n \to \infty$. On the other hand, the entry-wise restriction $a \in [0, 1]$ that holds for a mask results in a significant reduction in $\|d\|_2$. Due to this, in order to get an LMMSE that is bounded away from the trivial $\int d(x) dx$, one needs an exposure time that is $O(n)$. Of course, this is not surprising; there are strong benefits to lenses when they are available. The need for long exposure times for coded apertures is also a known phenomenon, consistent with the emphasis of [2] on "hypothesis tests" between scenes as opposed to resolving full detail.

One way to interpret increased $t$ is that it reduces noise relative to the signal. All our main results established in the sequel (eqs. [3] to [5]) show that one can construct apertures that are guaranteed to be tight within a constant factor of $t$. Under the above interpretation, what we establish rigorously is that our results are tight to within a universal constant number ($\approx 18.30$) of dB, regardless of the scene correlation structure given by $d$. This factor may be read off from $2M(n)^2$ of Prop. [3].

3. RESULTS

The goal of optimal aperture design (aka optimal $a$) is to minimize the LMMSE formula subject to the scene model, denoted as follows:

$$m^*(n, t, W, J, d, a) = \min_a m(n, t, W, J, d, a).$$

Let us first understand why the minimization above is a challenging problem. Consider the even simpler problem in which the optimal transmissivity, say $\rho_0$, is given to us. Then, although $a \in [0, 1], \rho(a) = \rho_0$ is a convex constraint, the LMMSE (1) which we wish to minimize is neither convex nor quasiconvex in $a$, since $1/(1 + cx^2)$ lacks any of these behaviors.

In order to solve this problem, our general approach is as follows. First, we use Parseval’s identity that relates time and frequency space. Under a fixed power budget, it is easy to solve for the optimal spectrum allocation $|\hat{a}|^2$ by studying the well-behaved convex $1/(1 + cx^2)$ that has a solution given by waterfilling [2]. Next, we are faced with the “coefficient problem” of finding a $a \in [0, 1]$ with given spectrum allocation. To address this, we appropriately apply a theorem of Nazarov [16, p. 5]. An exposition of Nazarov’s work together with the context he draws from (e.g., the geometric ideas of [20], along with the analytic ideas of [21]) may be found in [22].

3.1. Lower bound

We first derive a lower bound for LMMSE (1) based on waterfilling (see, e.g., [23, Thm 19.7]). For notational ease, we let $\gamma = t/(\Omega(W + J))$ throughout.

**Proposition 1.** Let $a$ satisfy $p(a) = \rho$. Then:

$$m(n, t, W, J, d, a) \geq \frac{1}{\gamma^2 + n^2 \rho^2} + \frac{1}{\sum_{i=0}^{n-1} d_i^2} + \frac{1}{\gamma P_t}. \quad (2)$$

Here $P_t = (1/\gamma)(T - 1/a_1)^+ \text{ and total power } P = \sum_{i=1}^{n-1} P_i = n(\rho + (n\rho - |\rho|^2)^2) - n^2 \rho^2$. Also note $P \leq n^2 \rho(1 - \rho)$. We remark that (2) is sharp if and only if $|\hat{a}|^2 = P_i$ for $0 < i < n$.

**Proof.** We have $\hat{a}_0 = n \rho$, giving the first term. For the nonzero frequencies, we use the fact that the maximum of $\sum_i x_i^2 \text{ subjected to } x_i \in [0, 1]$ and $\sum x_i = r$ is $r = |x| + (r - |x|)^2$. This, together with Parseval’s identity, yields an upper bound on the power of the nonzero frequencies. Waterfilling, modified to study $\frac{1}{1 + ax}$ as opposed to $\log(1 + ax)$, then gives the proposition. The floors are removed to get the $P$ upper bound by $x^2 \leq x$ for $0 \leq x \leq 1$.

Note that minimizing the right hand side over $p$ gives a lower bound on $m^*(n, t, W, J, d)$. This task is trivial numerically, but in general difficult analytically. We denote this optimal $\rho$ by $\rho^*$ henceforth.

3.2. Upper bound

Our goal here has been set from [2]. Conceptually, the design issue is finding a $a \in [0, 1]$ with prescribed lower bounds $|\hat{a}|^2 \geq P_i$. In general, this is impossible to do, and thus our lower bound [2] is not sharp in all settings. However, it should be noted that sharp cases do exist. Perhaps the conceptually simplest example is the analog of [2] for a lens, where our bound is sharp.

Our general approach is to simply step back by a factor $C$ and obtain a $a \in [0, 1]$ with $|\hat{a}|^2 \geq P_i/C$. What we do next is address how we can guarantee such a $C$. We shall move from simpler to
more complex situations, and accordingly start off with i.i.d. scenes where for infinitely many $n$ one does not need the full generality of Nazarov’s solution.

3.2.1. i.i.d. scenes

Recalling that $d(x) = \theta$ is constant, the waterfiling asks for a 0, 1 sequence with uniform spectrum allocation after the DC term (“spectrally flat sequences”). As already noted in [4] [15], one can certainly construct such spectrally flat sequences for infinitely many values of $n$, as long as they are “unbiased” with $\rho = 1/2 - o(1)$. This meets the lower bound (as $n \to \infty$) as long as the optimal $\rho$ is $1/2 - o(1)$ for the given $t, W, J$. A natural question is how good is using an “unbiased” spectrally flat sequence when $\rho \neq 1/2$? The answer is given in the following:

**Proposition 2.** Let $\theta, W, J$ be fixed and let $d = (\theta/n)1$. Then for infinitely many $n$, there exists $a \in [0, 1]$ such that:

$$m(n, 2t, W, J, d, a) \leq m^*(n, t, W, J, d).$$

In other words, “unbiased” spectrally flat sequences are always guaranteed to achieve optimal LMMSE at the expense of increasing the exposure time $t$ by a factor of at most 2. In the sequel, we show how one can reduce this factor even further.

This is achieved by using spectrally flat sequences with $\rho = 1/8 - o(1)$ and $\rho = 1/4 - o(1)$, and allows us to refine 2 to $8/7$. The construction of these is based on well-established cyclotomic number computations [22, Art. 350] in number theory: 1/4 corresponds to quartic residues [25], and 1/8 corresponds to octic residues [26]. It should be emphasized, however, that in contrast to the case in which $\rho = 1/2$, the existence of such sequences for infinitely many values of $n$ is not guaranteed, because no single-variable quadratic taking on infinitely many primes is known [27].

Of perhaps greater importance is the fact that the octic residue constructions of [25] rely upon primes that come from a second order linear recurrence with rather large coefficients, arising as the solutions of the Brahmagupta-Pell equations. There is thus a paucity of such constructions, indeed [26] gives only two such $n$ below $10^5$, namely $n = 73$ and $n = 26041$. On the other hand, the quartic residue constructions are reasonably numerous, with over 150 of them available below $10^7$. Even restricting ourselves to the quartic residues allows us to tighten from 2 to $4/3$. Summarizing all of the above, we have:

**Proposition 3.** Let $\theta, W, J$ be fixed and let $d = (\theta/n)1$. Then for some values of $n$ that exist even beyond, e.g., $10^5$, there exists $a \in [0, 1]$ such that:

$$m(n, (8/7)t, W, J, d, a) \leq m^*(n, t, W, J, d).$$

Moreover, for many ($> 150$ for $n < 10^7$) values of $n$ that exist even beyond, e.g., $10^9$, there exists $a \in [0, 1]$ such that:

$$m(n, (4/3)t, W, J, d, a) \leq m^*(n, t, W, J, d).$$

**Proof of Props.** The “difference sets” of [25] [26] are in our language spectrally flat sequences. The constant factor is given by the following single variable optimization. In view of [2], let $f_\alpha(\rho) = (\rho(1 - \rho))/((\alpha + \rho)$ defined on $[0, 1]$; $\alpha$ corresponds to $W/J$. The numerator comes from the power bound, the denominator from the noise penalty. Then, $M(\alpha, \rho) = (\sup_x f_\alpha(x))/f_\alpha(\rho)$ is the multiplicative loss factor for a fixed $W/J$ and fixed $\rho \in \{0.125, 0.25, 0.5\}$. One may then optimize over $\rho, a$ to get the constant $\rho_{\alpha}$. This proof, modified to $\rho \in \{0.25, 0.5\}$ and $\rho \in \{0.5\}$, also yields [45] and [3] respectively. The fact that there are infinitely many $n$ for $\rho = 0.5 - o(1)$ follows from the quadratic residue construction together with the well known fact that there are infinitely many primes $p = 4k + 3$ (see, e.g., [23, Chap 7]).

3.2.2. Correlated scenes

We now turn to correlated scenes. Here the waterfiling is nontrivial, and asks for an unequal spectrum allocation. We therefore invoke Nazarov’s solution to the coefficient problem [16, p. 5], and provide a statement here specialized to the DFT and $l_2(n)$ that we use.

First, some notation. Let us define inner products with respect to the uniform probability distribution on $\{0, 1, \ldots, n - 1\}$. Let $0 \leq i, j \leq n - 1$, and let $\psi_j$ be a orthonormal basis for the DFT on real sequences. Explicitly, let $h = [(n - 1)/2]$. Let $\psi_j(i) = 1$, $\psi_j(i) = \sqrt{2}\cos(\omega ji)$ for $0 < j < h$, $\psi_j(i) = \sqrt{2}\sin(\omega ji)$ for $h < j \leq n$. If $n$ is even, let $\psi_n(i) = \cos(\omega hi)$, otherwise $\psi_n(i) = \sqrt{2}\cos(\omega hi)$. Finally, let $\beta(n) = \min_i |\psi_j(i)|$.

**Theorem 1** (Nazarov). Let $M(n) = ((3\pi/2))\beta(n)^{-2}$. Let $0 \leq p_0, p_1, \ldots, p_{n-1}$ be such that $\sum p_j = 1$. Then there exists a $b \in [-M(n), M(n)]$ with $|\langle b, \psi_j \rangle|^2 \geq p_j$ for all $0 \leq j \leq n - 1$.

With this in hand, we are able to reach a far more general version of Prop. 4 valid for any $n$ and any scene prior $d$. Also, in Sec. 3.3, we show how to actually construct such tight sequences whose existence is guaranteed by [14].

**Proposition 4.** For all $n, t, W, J, d$, there exists a $a \in [0, 1]$ such that:

$$m(n, 2M(n)^2t, W, J, d, a) \leq m^*(n, t, W, J, d).$$

**Furthermore, we have:**

$$M(n) \in [(3\pi^3)/16 + o(1), 3\pi + o(1)].$$

The justification of the tightness of [3] lies in establishing [1], which we do first. The phenomenon is captured by the factorization of $n$, with the best, that is the largest, $\beta$ occurring for $n$ prime, and the worst occurring for $n$ divisible by 4. We have the following Lemma which establishes [3]:

**Lemma 1.** $\beta(n) \in \left[\frac{1}{\sqrt{2}} + o(1), \frac{2\sqrt{2}}{\pi} + o(1)\right]$ as $n \to \infty$. Moreover, if we restrict to $n$ being prime, $\beta(n) = \frac{2\sqrt{2}}{\pi} + o(1)$.

**Proof sketch.** We give a full proof for the $n = p$ prime case. Then, for any $j \neq 0$, $ij$ sweeps over $\{0, 1, \ldots, p - 1\}$, modulo $p$. Thus, really one is looking at a Riemann sum approximation to $\int_0^1 |\cos(2\pi x)| dx = 2/\pi$. The $l_2$ norm of $\cos(2\pi x)$ on $[0, 1]$ is $1/\sqrt{2}$, completing the prime case. The composite case is more involved, as it needs to take into account the divisor structure of $n$, which prevents such symmetry of the cosine vectors. Once accounted for, the natural idea is to use Euler-Maclaurin summation, with standard modifications by, e.g., mollifiers to take into account the lack of smoothness of $|\cos(x)|$ at its zeros. However, the mechanics are perhaps simplest in our specific setting when one uses short quadratic splines around the zeros to get a $C^1$ approximation of any desired accuracy to $|\cos(x)|$ while not changing the uniform derivative bound. We omit a full proof due to space constraints; see, e.g., [29] for the mechanics of how this is done in general.
Proof of Prop. 2 Thm. 1 with \( \rho_0 = 0 \) and \( p_j = P_j / \sum_j P_j \) for \( 0 < j < n \) yields a \( b \) with \( |b|_\infty \leq M(n) \) and \( (b, \psi_j)^2 \geq p_j \) for \( 0 < j < n \). Without loss, we may assume that \( (b, \psi_0)^2 \leq 0 \), else simply flip signs. Stitching the \( \psi_j \) back to complex exponentials and recalling the upper bound \( P \leq n^2 \rho(1 - \rho) \), this gives \( |b_j|^2 \geq P_j / (\rho(1 - \rho)) \). Consider \( a = (b + M(n)) / 2M(n) \). Then, \( a \in [0, 1], \rho(a) \leq 0.5 \), and \( |a_j|^2 \geq P_j / (4M(n)^2 \rho(1 - \rho)) \) for \( 0 < j < n \). We are now in a similar situation to that of Prop. 2 except with an extra \( M(n)^2 \) factor, and the fact that \( \rho(a) \leq 0.5 \) instead of \( \rho(a) = 0.5 + o(1) \). The latter is no problem, as lower \( \rho \) only helps us with the shot noise term, and the former simply multiplies the 2 of (3) by \( M(n)^2 \). 

3.3. Greedy algorithm

Here we propose a (heuristically) efficient algorithm to construct vectors \( a \) that satisfy the conditions of Prop. 4. This algorithm has its roots in Nazarov’s original proof. At a high level, Nazarov’s theoretical construction boils down to finding a “sign cortège” [14] p. 6] that is globally optimal for a certain real-valued Boolean function of \( n \) signs, taking exponential time in the worst case. However, a closer examination of Nazarov’s proof reveals that one simply needs a sign cortège that is locally optimal in the sense of Hamming geometry for the proof to work. Our observation suggests a natural greedy algorithm where one starts with a random cortège, and then flips one sign at a time if it improves the objective, repeating until no further improvement is possible. In our simulations this runs very fast. For example, on our standard laptop, we can generate apertures for \( n = 2000 \) in 4 seconds. This superficially resembles the situation of the simplex algorithm and the smoothed analysis of [30], or more directly recent work on max-cut [31]. Direct application of the methods of [31] to obtain theoretical guarantees runs into difficulties with the nonlinear change in objective with a single bit flip in our setting, unlike the linear change for max-cut. As such, we defer theoretical study of the greedy algorithm given here to future work.

3.4. Simulations

We give a simple illustration in Fig. 1 which confirms the following intuition based on our main results eqs. (3) to (5). With an i.i.d. scene prior, one would prefer using the spectrally flat construction as opposed to the one coming from Nazarov’s theorem due to the smaller constant. On the other hand, with a strong prior—e.g., a bandlimited one—the waterflling becomes highly skewed, and one would favor the one coming from Nazarov’s theorem as it takes into account such strong skewing of the desired spectrum. For completeness, we also include the performance of a random on-off sequence with density \( \rho \) [14], where \( \rho \) is optimized over \([0, 1]\) for each \( t \).

4. DISCUSSION AND FUTURE WORK

Our refined analysis of a model drawing heavily from [14] yields tight conclusions across all scene correlation patterns and noise regimes, with sharp conclusions available in some specific scenarios. Moreover, we give heuristically efficient algorithms for the generation of optimal coded apertures. We also note that similar conclusions to our main results eqs. (4) to (5) also hold for MI and Gaussian statistics of [14], simply because of the form of the expression for MI.

Furthermore, we note that our conclusions generalize naturally to 2D apertures, and in particular we have a tight characterization of optimal coded apertures in that setting. Concretely, one simply needs to take \( \beta(n)^2 \) as opposed to \( \beta(n) \) due to the squaring of the \( l_1 \) lower bound for the 2D DFT. The rest of the analysis of Thm. 1 and Prop. 2 carries over naturally, with the orthogonal basis provided by products of \( \psi_j \). We emphasize that this works regardless of the scene prior, even ones which are not separable. With an i.i.d. prior, separable apertures are optimal up to constants as in 1D, and in fact taking a product of spectrally flat apertures yields natural analogs of Props. 2 and 3. However, with other priors, it seems like one needs the generality provided by Thm. 1. This work thus also answers the question of 2D apertures raised in [14]. We also view experimental verification of these ideas as a worthwhile task.

As noted in [14, 9] raises the question of whether continuous-valued masks perform better than binary-valued ones. This work sheds some light on this: the solution of Nazarov which we have shown is tight does seem to use the flexibility of the \( l_\infty \) norm in an essential way; see, e.g., [32] p. 12] for more on this. And more specifically, we have numerical evidence for

\footnote{Code: https://github.com/gajjanag/apertures}
finite $n$; to give a concrete example, for $n = 13$, the mask 
$[1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0]$ has optimal LMMSE for an i.i.d. 
scene over binary-valued masks for $\rho = 6/13, \theta = 0.01, W = J = 
0.001, t = 130$, but is improved upon by the continuous-valued 
mask whose first entry is equal to $\epsilon$ and whose $i$th entry is equal to 
$1 - \epsilon/6$ if $i - 1$ is a quadratic residue modulo 13, and 0 otherwise, 
for $0.26 \leq \epsilon \leq 0.34$.

Although Prop. 4 shows universal tightness across all priors, 
even “extreme” ones like bandlimited ones, the constant is worse 
than that for a spectrally flat construction for i.i.d. scenes. The better 
performance of spectrally flat constructions over the ones inspired 
by Nazarov’s theorem seems to extend to other “natural” priors like 
the $f^{-\gamma}$ one, as the waterfilling still yields something that is nearly 
“flat”. It might be interesting to quantify and understand the “flat-
ness” of the waterfilling for “natural” priors.

One issue that we have not addressed here or in [14] is the equal 
scaling of $n$ at both sensor and scene. One natural way to address 
this is letting $A$ be $m \times n$, or alternatively one could study a con-
tinuous model. Another issue is obtaining a good understanding of 
mask/lens combinations. This will require not only updates to the 
simple propagation model studied here and in [14], but also a refined 
understanding of the cost tradeoffs between lenses and apertures.

Stepping back from imaging problems, one may ask the question 
of where else Nazarov’s theorem can be used in applied contexts, 
something also raised implicitly in [17]. For example, as Nazarov’s 
theorem does not care about orthogonality, but merely a

5. ACKNOWLEDGEMENT

Ganesh Ajjanagadde thanks Prof. Henry Cohn for discussions on pseudorandomness.

6. REFERENCES


