2005 Antenna Applications Symposium, Robert Allerton Park, Monticello, IL, Sep. 21–23: http://www.ecs.umass.edu/ece/allerton/

EFFICIENT SUBARRAY REALIZATION THROUGH LAYERING

J. O. Coleman* K. R. McPhail P. E. Cahill D. P. Scholnik

Naval Research Laboratory (http://www.nrl.navy.mil/), Radar Division
Code 5328, Signal Processing Theory & Methods Section

Abstract: In many "digital" planar receive arrays proposed today element outputs are not digitized directly. Instead, the outputs of a lesser number of overlapped identical analog subarrays are digitized. Here the severe cost/performance tradeoff usually encountered when choosing subarray sizes and overlaps is dramatically improved by replacing the usual single layer of analog subarray processing with multiple layers of subarray processing, each further reducing the spatial density of signal outputs to be processed. The set of signals digitized can be made quite sparse while realizing most of the required large subarray overlap in the low-cost intermediate layers. Here we develop the theory and present a major design example in which the cost/performance tradeoff is improved a full order of magnitude by one key measure. We discuss receive arrays, but the ideas apply to transmit arrays as well.

1. Introduction

For cost reasons, large receive arrays generally use digital beamshaping not on the array of element outputs directly but only on a less-dense array of analog subarray outputs, each created by analog hardware that combines signals from several elements. The array community is currently using subarray terminology for two fundamentally different approaches. One of them simply splits the beamshaping sum on the left into subsums as on the right:

$$\sum_{n \in \substack{\text{element} \\ \text{identifiers}}} w_n \, X_n(f) = \sum_{k \in \substack{\text{subarray} \\ \text{identifiers}}} \; \sum_{n \in \substack{\text{identifiers} \\ \text{of elements in} \\ \text{subarray } k}} w_n \, X_n(f).$$

Sometimes a complex normalization amplitude is factored out of the inner sum and included in the analog processing, resulting in

$$\sum_{k \in \substack{\text{subarray} \\ \text{identifiers}}} \; c_k \sum_{n \in \substack{\text{identifiers} \\ \text{of elements in} \\ \text{subarray } k}} w'_n \, X_n(f),$$

but either way this architecture is only a routine partitioning of operations between analog processing—the inner sum—and digital processing—the outer sum. Neither the weights

* This work was supported by the Office of Naval Research (http://www.onr.navy.mil/) and by base Science & Technology funds at the Naval Research Laboratory (http://www.nrl.navy.mil/). Author contacts: http://alum.mit.edu/www/jeffc, *kelly.mcphail@nrl.navy.mil, pcahill@alumni.upenn.edu, dan.scholnik@nrl.navy.mil*

within a subarray nor the normalizing coefficients c_k have separate meaning, so this is subarraying in hardware terms only. From a signal-processing viewpoint the scheme is uninteresting (and the subarray terminology inappropriate). We will not refer to it further.

In the other subarraying approach, our focus here, subarray sums use identical weights, so in the array-pattern analysis a subarray array factor can be factored out of the sum over subarray outputs. The overall array pattern is then product of an element pattern and separate array factors for the analog subarray combining and for the final digital combining. (See (40) for the mathematics.) Here we extend this to multiple layers of subarray combining, resulting in more than two array factors in the final pattern. Conventionally only the last stage is digital, but in principle A/D conversion can take place between any two processing layers or not at all. We focus on the narrowband case, but the wideband extension should not be difficult.

Using one subarray-combining layer or several, the subarray approach results in ultimate, last-step combining of signals whose spatial density is much lower than that of the elements themselves. The periodic duplications of the "main beam" of the associated ultimate array factor are closely spaced and would become grating lobes in the overall array pattern except for the action of the penultimate and earlier array factors associated with subarray combining, the function of which is precisely to suppress these repeated beams. Conventionally, one subarray array factor must suppress all these potential grating lobes, but in the multi-layer approach here, that suppression task is split among several simpler array factors. For an example, look ahead to Fig. 6. There the product of the three individual array factors on the upper left is the overall array factor on the lower right. This layered approach lowers costs substantially, because each subarray-processing layer reduces the spatial density of signals processed by following stages. Much of the subarray processing is performed at less-expensive lower spatial densities.

The core ideas of this multi-layer subarray approach are adapted from ideas routinely used in digital signal processing [1] to realize the functionality of severe filtering and high-ratio decimation using several stages of simple filtering and low-ratio decimation. The conventional DSP use of stopbands of earlier filtering stages to suppress periodically repeated passbands of later filtering stages is exactly analogous here to using array factors of earlier subarray-combining stages to suppress periodically repeated beams of the array factors of later stages, the would-be grating lobes.

In the sequel we do not assume the reader to be familiar with either this earlier DSP approach or with the single-layer subarray approach. We simply develop what ideas we need in standalone fashion, from the beginning. We avoid commonly made mathematical simplifications—a separable array factor $A(x, y) = A_x(x)A_y(y)$ and simple uniform element spacing in x and y are the obvious ones—when those simplifications would obviously increase hardware cost significantly in large array applications. (Most large arrays are for expensive government-owned radar systems, so we dedicate this work to the taxpayers.) As a result, this paper is not particularly light reading. It is a theoretical paper and includes as much of the mathematics of arrays as will reasonably fit within conference page limits. This means, in particular, that it includes a good bit of basic lattice theory, developed

here in a specific context: 2D lattices in a 3D space. We have not centralized the lattice theory in one place but intersperse it throughout the paper, introducing topics as they are needed. Most of the material introduced here on lattices can be found in the first few pages of Conway and Sloane [2], the bible of lattice theory, and in the first few dozen pages of the group-theory section of a good senior-level abstract-algebra text like Herstein [3].

2. Background: A Signal-Processing View of Arrays and Subarrays

This section develops, as background, the theory of the linear receive array—linear processing, not linear spacing—and of its realization with a conventional single-level subarray architecture. This establishes not only notation but a detailed analytical framework in which the architecture of the multi-layer subarray will be subsequently developed in Section 3.

The mathematical approach here is a little unconventional, because we are attempting to simultaneously address readers working in electromagnetics and in signal processing. Our intent is to require prior familiarity in each of those areas only with material that should be familiar to any researcher in electrical engineering. As a result, either community may find some material to be simple bordering on tedious while other material is somewhat challenging. Our bias is clear: we work in signal processing, and this paper addresses a traditional electromagnetics topic with signal-processing tools. The view of electromagnetics here is a signal-processing view.

2.1. Electromagnetic Fields

We rely on familiar Fourier-transform properties later, particularly symmetry and convolution properties, so for frequency-domain work we will use Fourier transforms of electromagnetic field variables in 2D, 3D, and 4D and not the superficially similar complex fields common in electromagnetics. So here we establish notation and relate our transformed fields to those conventional complex-field representations. We begin with the real electric-field vector $\vec{\mathbf{e}}(\mathbf{t})$ as a function of length-dimensioned 4D spacetime vector $\mathbf{t} \triangleq \left[\begin{smallmatrix}\mathbf{x}\\ ct\end{smallmatrix}\right]$. We then express it as this Fourier integral on a 4D spacetime frequency $\mathbf{f} \triangleq \big[\mathbf{k}, f/c\big]$, where $d\mathbf{f}$ is differential 4D volume:

$$\vec{\mathbf{e}}(\mathbf{t}) = \int \overrightarrow{\mathbf{E}}(\mathbf{f})\, \mathrm{e}^{\mathrm{j}2\pi \mathbf{f}\cdot\mathbf{t}}\, d\mathbf{f}. \tag{1}$$

The wave parameterization here differs from those common in the electromagnetics community. When electromagnetic source variables are of form $\mathrm{Re}\{A\mathrm{e}^{\mathrm{j}\omega t}\}$ with various complex amplitudes A and a common $\omega > 0$, Maxwell's equations require the electric field in source-free regions to comprise components of form $\mathrm{Re}\{\overrightarrow{\mathbf{E}}_0\, \mathrm{e}^{\mathrm{j}(\omega t - \boldsymbol{\kappa}\cdot\mathbf{x})}\}$, which is parameterized by complex vector amplitude $\overrightarrow{\mathbf{E}}_0$ and complex wavenumber vector $\boldsymbol{\kappa}$ with the latter governed by *Helmholz relation* $(\boldsymbol{\kappa}c/\omega)\ \cdot\ (\boldsymbol{\kappa}c/\omega) = 1$ and where the dot product is non-Hermitian: in Cartesian coordinates it is simply a sum of coordinate products with no conjugation. Using real vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to write $\boldsymbol{\kappa}c/\omega = \boldsymbol{\alpha} - \mathrm{j}\boldsymbol{\beta}$ reduces that relation to $\boldsymbol{\alpha}\cdot\boldsymbol{\beta} = 0$ and $\|\boldsymbol{\alpha}\|^2 = \|\boldsymbol{\beta}\|^2 + 1$ and this component with a single complex wavenumber

to

$$\mathrm{Re}\big\{\,\overrightarrow{\mathbf{E}}_0\,\mathrm{e}^{\mathrm{j}(\omega t-\boldsymbol{\kappa}\cdot\mathbf{x})}\big\}=\mathrm{e}^{-\omega\boldsymbol{\beta}\cdot\mathbf{x}/c}\,\mathrm{Re}\big\{\overrightarrow{\mathbf{E}}_0\,\mathrm{e}^{\mathrm{j}\omega(t-\boldsymbol{\alpha}\cdot\mathbf{x}/c)}\big\},\tag{2}$$

which has loci of fixed-phase points $\mathbf{x}$ moving in the direction of vector $\boldsymbol{\alpha}$ at speed $c/\|\boldsymbol{\alpha}\|$ and therefore with wavelength $\lambda/\|\boldsymbol{\alpha}\|$, where $\lambda\triangleq 2\pi c/\omega$.

If $\|\boldsymbol{\alpha}\|=1$ so that $\boldsymbol{\beta}=0$ makes wavenumber $\boldsymbol{\kappa}$ real, component (2) is a propagating plane wave represented exactly by Fourier integral (1) with

$$\overrightarrow{\mathbf{E}}(\mathbf{f})=\tfrac{1}{2}\overrightarrow{\mathbf{E}}_0\,\delta\big(\mathbf{f}-\tfrac{1}{2\pi}\big[-\boldsymbol{\kappa},\omega/c\big]\big)+\tfrac{1}{2}\,\overrightarrow{\mathbf{E}}_0^{\,*}\,\delta\big(\mathbf{f}-\tfrac{1}{2\pi}\big[\boldsymbol{\kappa},-\omega/c\big]\big),$$

a conjugate pair of Fourier components at 4D frequencies that are negatives of each other. Propagation is at speed c, and if $\mathbf{f}=\big[\mathbf{k},\,f/c\,\big]$ is either impulse location, wavelength $\lambda=c/|f|$ for $f\neq 0$. Constancy of $\mathrm{j}2\pi\mathbf{f}\cdot\mathbf{t}$ in (1) implies propagation in the $-(c/f)\,\mathbf{k}$ direction: in the $\mathbf{k}$ direction when $f<0$ and in the $-\mathbf{k}$ direction when $f>0$. Wavenumber $\boldsymbol{\kappa}=-2\pi\,\mathrm{sgn}(f)\,\mathbf{k}$, so $f>0$ puts $\mathbf{k}$ and $\boldsymbol{\kappa}$ in opposite directions, and we can think of the direction of $\mathbf{k}$ as the *look direction*. The Helmholz relation specializes here to $\|\boldsymbol{\kappa}\|=\omega/c$ or $\|\mathbf{k}\|=|f|/c$ or even, letting $\widehat{k}$ be a look-direction unit vector, $\mathbf{k}=\widehat{k}\,f/c$, so impulse frequencies are in set $\{\mathbf{f}:\,\|\mathbf{k}\|=|f|/c\}$, the *4D Helmholz cone*. (A set closed under scaling by positive constants is a cone.) A vector $\mathbf{k}$ for which $\|\boldsymbol{\kappa}\|\lessgtr\omega/c$ or $\|\mathbf{k}\|\lessgtr|f|/c$ is $\overset{\textit{inside}}{\textit{outside}}$ the 4D Helmholz cone.

If $\|\boldsymbol{\alpha}\|>1$ instead so that $\|\boldsymbol{\beta}\|>0$, then field component (2) is an evanescent wave with an amplitude that decays exponentially at rate $\|\omega\boldsymbol{\beta}/c\|$ in the direction of vector $\boldsymbol{\beta}$, orthogonal to the direction of $\boldsymbol{\alpha}$ in which phase propagates. The quantity $\mathrm{Re}\{$ *stuff* $\}$ in (2) has 4D Fourier transform

$$\tfrac{1}{2}\overrightarrow{\mathbf{E}}_0\,\delta\big(\mathbf{f}-\tfrac{1}{2\pi}\tfrac{\omega}{c}\big[-\boldsymbol{\alpha},1\big]\big)+\tfrac{1}{2}\,\overrightarrow{\mathbf{E}}_0^{\,*}\,\delta\big(\mathbf{f}-\tfrac{1}{2\pi}\tfrac{\omega}{c}\big[\boldsymbol{\alpha},-1\big]\big),\tag{3}$$

in which 4D impulse frequencies $\mathbf{f}$ have $\|\mathbf{k}\|=\|\boldsymbol{\alpha}\|\,|f|/c>|f|/c$. Propagation is at speed less than c and with wavelength less than $|f|/c$, and the decay-rate vector $(\omega/c)\boldsymbol{\beta}$ lies on a circle of radius $(\omega/c)\sqrt{\|\alpha\|^2-1}=2\pi\sqrt{\|\mathbf{k}\|^2-|f|^2/c}$ in the plane normal to $\boldsymbol{\alpha}$. The spatially varying amplitude in (2) is not suitable for Fourier analysis as written, as it disguises the fact that boundary conditions must limit the spatial extent of its applicability. Physical fields can't grow exponentially in the $-\boldsymbol{\beta}$ direction forever.

Consider a simple case. Use a scalar β and a unit vector $\widehat{\beta}$ to write $\boldsymbol{\beta}=\beta\widehat{\beta}$, and suppose boundary conditions require $\overrightarrow{\mathbf{e}}_0(\mathbf{t})=0$ wherever $\mathbf{x}\cdot\widehat{\beta}<x_b$. Using a unit step function $U(x)$, the truncated evanescent field then is just

$$U(\mathbf{x}\cdot\widehat{\beta}-x_b)\,\mathrm{e}^{-\omega\boldsymbol{\beta}\cdot\mathbf{x}/c}\,\mathrm{Re}\big\{\overrightarrow{\mathbf{E}}_0\,\mathrm{e}^{\mathrm{j}\omega(t-\boldsymbol{\alpha}\cdot\mathbf{x}/c)}\big\}.\tag{4}$$

Decomposition $\mathbf{x}=x\widehat{\beta}+\underline{\mathbf{x}}$, where $\underline{\mathbf{x}}\perp\widehat{\beta}$, makes the amplitude factor just $\mathrm{e}^{-\omega\beta x_b/c}\,U(x-x_b)\,\mathrm{e}^{-\omega\beta(x-x_b)/c}$, the usual impulse response of a first-order lowpass filter with an additional scale factor and an offset in x. Using $\mathbf{x}=k\widehat{\beta}+\underline{\mathbf{k}}$ with $\widehat{\beta}\perp\underline{\mathbf{k}}$, this factor has 4D Fourier transform

$$\mathrm{e}^{-\omega\beta x_b/c}\,\frac{\mathrm{e}^{-\mathrm{j}2\pi k x_b}}{\mathrm{j}2\pi k+\omega\beta/c}\,\delta(\underline{\mathbf{k}})\,\delta(f/c),\tag{5}$$

the support of which is the entire 4D line along unit vector $\big[\widehat{\beta}, 0\big]$. The Fourier transform of product (4), the convolution of the transforms (5) and (3) of its two factors, is then supported on the parallel line translates

$$\mathbb{R}\big[\boldsymbol{\beta}, 0\big] \pm \tfrac{1}{2\pi}\tfrac{\omega}{c}\big[\boldsymbol{\alpha}, -1\big], \tag{6}$$

which lie entirely outside the 4D Helmholz cone.

When boundary conditions further restrict the fields, (4) is effectively multiplied by some $\{0, 1\}$-valued masking function $m(\mathbf{t})$, and convolution with its transform $\mathbf{M}(\mathbf{f})$ then broadens support (6), usually to all of $\mathbf{f}$ space.

2.2. Receive Antennas and Arrays

The map from source variables to electromagnetic fields is linear with a trivial nullspace—nonzero source variables imply nonzero fields—and therefore has a linear inverse. If we call the fields in the absence of the receiver antenna the "pretend fields," there is therefore a linear map, that inverse, from the pretend fields to the nonlocal source variables, another from there on to the actual fields that obtain when the antenna is present, and a third from the latter fields to the actual antenna output variables. The last is of course because currents and voltages are just line integrals of fields. The composition of these three linear maps is a linear map from the pretend fields to the antenna output variables. We can quite reasonably suppose that the output variables depend only on values of the pretend fields in the vicinity of the antenna, and if this antenna-local region is in the far field of the sources it is enough to know either the pretend electric field or the pretend magnetic field, as there each can be determined linearly from the other.

So let $\vec{\mathbf{e}}(\mathbf{t})$ be the pretend electric field—we drop the "pretend" designation hereinafter—created in the vicinity of the antenna by the sources of interest, and assume they are sufficiently distant that their evanescent fields are essentially zero. It follows that $\vec{\mathbf{E}}(\mathbf{f})$ in Fourier representation (1) represents plane-wave components only and that its support is confined to the 4D Helmholz cone. The $t=0$ output of a receive antenna nominally located at $\mathbf{x}=0$ is then a linear functional of $\vec{\mathbf{e}}(\mathbf{t})$ and so by various Riesz representation theorems [4] is characterized by a 4D vector $\vec{\boldsymbol{\phi}}$ of signed measures as

$$\int \vec{\mathbf{e}}(-\mathbf{t}) \cdot d\vec{\boldsymbol{\phi}}(\mathbf{t}), \tag{7}$$

where the minus sign is simply convenient. Now move the antenna to an arbitrary location $\mathbf{x}$ and denote its signal output at time t with $a\big(\big[\mathbf{x}, ct\big]\big)$. The $a\big(\big[\mathbf{x}, ct\big]\big)$ that results from field $\vec{\mathbf{e}}(\mathbf{t})$ equals the $a(0)$ that results from field $\vec{\mathbf{e}}(\mathbf{t} + \big[\mathbf{x}, ct\big])$, a spacetime-invariance property, so from (7),

$$a\big(\big[\mathbf{x}, ct\big]\big) = \int \vec{\mathbf{e}}\big(\big[\mathbf{x}, ct\big] - \mathbf{t}\big) \cdot d\vec{\boldsymbol{\phi}}(\mathbf{t}).$$

In terms of *antenna pattern* $\overrightarrow{\mathbf{\Phi}}(\mathbf{f})$ this 4D convolution is just Fourier integral

$$a\big([\mathbf{x}, ct]\big) = \int \overrightarrow{\mathbf{E}}(\mathbf{f}) \cdot \overrightarrow{\mathbf{\Phi}}(\mathbf{f})\ \mathrm{e}^{\mathrm{j}2\pi\mathbf{f}\cdot\mathbf{t}}\, d\mathbf{f}, \tag{8}$$

with $\overrightarrow{\mathbf{\Phi}}(\mathbf{f})$ taken as fundamental or derived as Fourier transform

$$\overrightarrow{\mathbf{\Phi}}(\mathbf{f}) = \int \mathrm{e}^{-\mathrm{j}2\pi\mathbf{f}\cdot\mathbf{t}}\, d\overrightarrow{\boldsymbol{\phi}}(\mathbf{t}).$$

It is more conventional to write this quantity $\overrightarrow{\mathbf{\Phi}}(\mathbf{f})$ as the product of a dimensionless vector gain and a dimensioned scalar, with the latter representing an "isotropic" antenna, and that vector gain is typically termed the pattern. We keep it all in $\overrightarrow{\mathbf{\Phi}}(\mathbf{f})$, however, to keep the notation simple.

A planar antenna with a hemispherical view. Suppose the antenna characterized by $\overrightarrow{\mathbf{\Phi}}(\mathbf{f})$ can be relocated at will to any point $\underline{\mathbf{x}}$ on the *array plane* normal to some *boresight-direction* unit vector $\widehat{b}$. If $\mathbf{x} = \underline{\mathbf{x}}$ and $\mathbf{k} = \underline{\mathbf{k}} + k\widehat{b}$ with $\underline{\mathbf{x}}$ and $\underline{\mathbf{k}}$ both in the array plane, (8) becomes

$$a\big([\underline{\mathbf{x}}, ct]\big) = \iint \left[\int \overrightarrow{\mathbf{E}}(\mathbf{f}) \cdot \overrightarrow{\mathbf{\Phi}}(\mathbf{f})\, dk \right] \mathrm{e}^{\mathrm{j}2\pi(\underline{\mathbf{k}}\cdot\underline{\mathbf{x}} + ft)}\ \frac{df}{c}\, d\underline{\mathbf{k}} \tag{9}$$

with $d\underline{\mathbf{k}}$ differential area in the array plane. Suppose no waves arrive from the hemisphere opposite boresight. Then $\overrightarrow{\mathbf{E}}(\mathbf{f}) = 0$ when $f\mathbf{k}\cdot\widehat{b} = fk < 0$. In the far field $\overrightarrow{\mathbf{E}}(\mathbf{f})$ is concentrated on the 4D Helmholz cone, so in fact $\overrightarrow{\mathbf{E}}(\mathbf{f}) = 0$ except where $\underline{\mathbf{k}}$ is in the *visible region* in the array plane defined by $\|\underline{\mathbf{k}}\| \le |f/c|$ and where $k = k_H(\underline{\mathbf{k}})$, the Helmholz k given by

$$k_H(\underline{\mathbf{k}}) = f/c\ \sqrt{1 - \|(c/f)\underline{\mathbf{k}}\|^2}. \tag{10}$$

In (9) then, $\overrightarrow{\mathbf{\Phi}}(\mathbf{f})/c$ can be replaced with *array-plane antenna pattern*

$$\underline{\overrightarrow{\mathbf{\Phi}}}(\underline{\mathbf{k}}, f) \triangleq \begin{cases} \overrightarrow{\mathbf{\Phi}}\Big(\big[\underline{\mathbf{k}} + k_H(\underline{\mathbf{k}})\widehat{b}, f/c\big]\Big)/c & \text{for } \underline{\mathbf{k}} \text{ in the visible region} \\ 0 & \text{for } \underline{\mathbf{k}} \text{ outside the visible region} \end{cases}$$

to obtain 1D and 2D Fourier-pair relationships

$$a\big([\underline{\mathbf{x}}, ct]\big) = \int A(\underline{\mathbf{x}}, f)\, \mathrm{e}^{\mathrm{j}2\pi ft}\, df \tag{11}$$

$$A(\underline{\mathbf{x}}, f) \triangleq \int \mathrm{A}(\underline{\mathbf{k}}, f)\ \mathrm{e}^{\mathrm{j}2\pi\underline{\mathbf{k}}\cdot\underline{\mathbf{x}}}\ d\underline{\mathbf{k}} \tag{12}$$

$$\mathrm{A}(\underline{\mathbf{k}}, f) = \left[\int \overrightarrow{\mathbf{E}}\Big(\big[\underline{\mathbf{k}} + k\widehat{b}, f/c\big]\Big)\, dk \right] \cdot \underline{\overrightarrow{\mathbf{\Phi}}}(\underline{\mathbf{k}}, f). \tag{13}$$

In (13) the concentration of $\overrightarrow{\mathbf{E}}(\mathbf{f})$ on the Helmholz cone makes the integral see only an impulse $\delta(k - k_H(\underline{\mathbf{k}}))$ times some vector-valued area to be extracted. The dot product

then applies frequency-selective antenna pattern $\underline{\overrightarrow{\mathbf{\Phi}}}(\underline{\mathbf{k}}, f)$ to determine the antenna output, represented in the temporal frequency domain by spatial Fourier pair $A(\underline{\mathbf{x}}, f) \leftrightarrow \mathrm{A}(\underline{\mathbf{k}}, f)$. Pattern $\underline{\overrightarrow{\mathbf{\Phi}}}(\underline{\mathbf{k}}, f)$ is *polarized* because it is vector valued and applied through a dot product.

Here we assumed the planar antenna was moved to location $\underline{\mathbf{x}}$ to obtain output $A(\underline{\mathbf{x}}, f)$. This is practical physically in one key case, considered next.

An infinite planar array: spatial lattice sampling. Let $\mathbf{b}_1$ and $\mathbf{b}_2$ be linearly independent spatial *basis vectors*, let $\mathbf{B} \triangleq \big[\mathbf{b}_1, \mathbf{b}_2\big]$ be a row vector—in this paper it is not a matrix—of these two basis vectors, and let $\mathbf{B}\mathbb{Z}^2$ be the set of points of the form $\mathbf{B}\boldsymbol{n} = \mathbf{b}_1 n_1 + \mathbf{b}_2 n_2$, where $\boldsymbol{n} = \left[\begin{smallmatrix} n_1 \\ n_2 \end{smallmatrix}\right]$ is an arbitrary two-vector of integers, expressed $\boldsymbol{n} \in \mathbb{Z}^2$. We say *lattice basis* $\mathbf{B}$ *generates* 2D *lattice* $\mathbf{B}\mathbb{Z}^2$. If $\mathbf{b}_1$ and $\mathbf{b}_2$ are in the array plane then so is their span $\mathbf{B}\mathbb{R}^2$ and lattice $\mathbf{B}\mathbb{Z}^2$. All lattices in this paper are in the array plane.

Move our hemispherical-view planar antenna to point $\underline{\mathbf{x}} = \mathbf{B}\boldsymbol{n}$ of this lattice. Its output is then $a\big(\big[\mathbf{B}\boldsymbol{n}, ct\big]\big)$, or in the frequency domain, by (12),

$$A(\mathbf{B}\boldsymbol{n}, f) = \left[\int \mathrm{A}(\underline{\mathbf{k}}, f)\, \mathrm{e}^{\mathrm{j}2\pi \underline{\mathbf{k}}\cdot\underline{\mathbf{x}}}\, d\underline{\mathbf{k}}\right]_{\underline{\mathbf{x}}=\mathbf{B}\boldsymbol{n}}. \tag{14}$$

We follow convention and term 1D Fourier pair $a\big(\big[\mathbf{B}\boldsymbol{n}, ct\big]\big) \leftrightarrow A(\mathbf{B}\boldsymbol{n}, f)$ the output of *element* $\boldsymbol{n} \in \mathbb{Z}^2$. The element outputs are simultaneously available only if the antenna is at all translations $\mathbf{B}\mathbb{Z}^2$ simultaneously, so the antenna structure must be periodic, invariant to translation by points of $\mathbf{B}\mathbb{Z}^2$.

What does it mean for an antenna to be periodic in structure in this way? A *period* of a lattice $\mathbf{B}\mathbb{Z}^2$ is any subset of array plane $\mathbf{B}\mathbb{R}^2$ that exactly tiles the plane when translated by all the lattice points. The period is not unique—artist M. C. Escher demonstrated that the period can be chosen creatively—but in terms of the 2×2 *Gram matrix* $\mathbf{G} \triangleq \mathbf{B}^T \cdot \mathbf{B}$ of the lattice basis, the pairwise dot products of the basis vectors, the area of one period is necessarily $\sqrt{\det(\mathbf{B}^T \cdot \mathbf{B})}$. We can write this as $|\mathbf{G}|^{1/2}$ using notation $|\mathbf{A}| \triangleq |\det(A)|$ for an arbitrary matrix $\mathbf{A}$. A physical array antenna is typically considered to be an infinite array of elements paired one-to-one with translated lattice periods, with the physical area of an element being no larger than the area of the period. Imagine such a physical array antenna, extending off to infinity, with a terminated feedpoint associated to (but not necessarily at) each lattice point. Here we refer to this entire, infinite structure as the antenna at location $\mathbf{B}\boldsymbol{z}$ for any $\boldsymbol{z} \in \mathbb{Z}^2$ we choose. Because this structure is periodic, we don't need to change anything physically to "move" the antenna to location $\mathbf{B}\boldsymbol{z}'$. We just observe a different feedpoint voltage or current.

This unconventional construction establishes that array-plane antenna pattern $\underline{\overrightarrow{\mathbf{\Phi}}}(\underline{\mathbf{k}}, f)$ governs what is seen at a feedpoint when the entire array antenna is physically present. It is of course never correct to measure in isolation the pattern of one period's worth of structure, one element in the conventional view, without the rest of the structure around it, and then pretend that it can be used straightforwardly in array calculations. It cannot. From our point of view that so-called "element pattern" is irrelevant, and no mention of this term

in this paper will refer to such a quantity. The pattern that matters here is that of the entire array of elements, but with output taken from one feedpoint only. We have eliminated the so-called problem of mutual coupling by simply building it into the pattern $\underline{\overrightarrow{\mathbf{\Phi}}}(\underline{\mathbf{k}}, f)$ of an infinite, periodic structure. In this lattice-sampling context we'll term $\underline{\overrightarrow{\mathbf{\Phi}}}(\underline{\mathbf{k}}, f)$ the *periodic element pattern* or just the *element pattern*.

We will speak of element $\boldsymbol{n}$ being *located at* $\mathbf{B}\boldsymbol{n}$, the antenna translation with which it is associated. This nominal feedpoint location will be the physical feedpoint location if the periodic antenna structure has a physical feedpoint at $\mathbf{x}=0$, but otherwise there will always be an offset, independent of $\boldsymbol{n}$, between nominal and physical feedpoint locations.

Two specific array lattices. Typically only two basic configurations of a boresight direction and a lattice basis are even candidates for array configurations. Here $\widehat{i}$, $\widehat{j}$, and $\widehat{k}$ are pairwise-orthogonal unit vectors in space, and $\lambda_u \triangleq c/|f_u|$ with frequency $|f_u|$ the upper edge of the (presumably narrow) signal band.

name	*boresight* $\widehat{b}$	*lattice basis*	*Gram matrix* $\mathbf{B}^T \cdot \mathbf{B}$	*element spacing*
rectangular array	$\widehat{k}$	$\mathbf{B}_\square \triangleq \frac{\lambda_u}{2}\left[\,\widehat{i},\ \widehat{j}\,\right]$	$\frac{\lambda_u^2}{4}\mathbf{I}$	$\frac{\lambda_u}{2}$
triangular array	$\frac{1}{\sqrt{3}}(\widehat{i}+\widehat{j}+\widehat{k})$	$\mathbf{B}_\triangle \triangleq \frac{\lambda_u}{\sqrt{6}}\left[\,\widehat{j}-\widehat{i}, \widehat{k}-\widehat{i}\,\right]$	$\frac{\lambda_u^2}{6}\left[\begin{smallmatrix}2&1\\1&2\end{smallmatrix}\right]$	$\frac{\lambda_u}{\sqrt{3}}$

By *element spacing* here we mean the smallest nonzero distance between lattice points, more generally termed the *nearest-neighbor distance* of a lattice. Lattices $\mathbf{B}_\square\mathbb{Z}^2$ and $\mathbf{B}_\triangle\mathbb{Z}^2$ are sometimes termed *square* and *hexagonal* lattices respectively. When we set $\mathbf{B}$ to either $\mathbf{B}_\square$ or $\mathbf{B}_\triangle$ we'll implicitly set boresight direction $\widehat{b}$ and Gram matrix $\mathbf{G}$ as indicated here as well. Here $\mathbf{B}_\square$ is included for completeness. Nearly all our specific results are based on $\mathbf{B}_\triangle$, which is superior in many ways.

2.3. Conventional Linear Array Processing

There is no hyphen above, so "Linear Array Processing" defaults grammatically to "Linear (Array Processing)" and does not mean "(Linear Array) Processing." We use no 1D, uniformly spaced, misnamed "linear" arrays.

Some signal-processing notation and basics. We need notation for signal sets at the inputs and outputs of subarray processing layers. Let $s^i_{\boldsymbol{n}}(f)$ denote member $\boldsymbol{n}\in\mathbb{Z}^2$ of the signal set identified by i. For brevity, we use

$$\begin{aligned} s^i_{\boldsymbol{n}} \quad &\text{for} \quad s^i_{\boldsymbol{n}}(f), \\ s^i \quad &\text{for} \quad \{s^i_{\boldsymbol{n}} : \boldsymbol{n}\in\mathbb{Z}^2\}. \end{aligned}$$

We define the Fourier transform of $s^i_{\boldsymbol{n}}$ on 2D index $\boldsymbol{n}$ in the usual way to be

$$S^i(\boldsymbol{f}) \triangleq \sum_{\boldsymbol{n}} s^i_{\boldsymbol{n}}\, \mathrm{e}^{-\mathrm{j}2\pi \boldsymbol{f}\boldsymbol{n}}, \tag{15}$$

with $\boldsymbol{f}$ any real two-element row vector. We sometimes write this $s^i \leftrightarrow S^i$.

We relate layers to each other through very few operations. First is *convolution* in index space. Let $g^i_{\boldsymbol{n}}$ denote member $\boldsymbol{n} \in \mathbb{Z}^2$ of the coefficient set—actually a function of $\boldsymbol{n}$—identified by i. Technically, indexing $g^i_{\boldsymbol{n}}$ on $\mathbb{Z}^2$ makes coefficient set g^i infinite, but when it matters we'll assume only some finite number $|g^i|$ of the coefficients can be nonzero. Set identifier i may be omitted when stating general principles. For convolution we use

$$\begin{gathered} \sum_{\boldsymbol{k}} \quad \text{for} \quad \sum_{\boldsymbol{k} \in \mathbb{Z}^2}, \\ s^j = g \star s^i = s^i \star g \quad \text{for} \quad s^j_{\boldsymbol{n}} = \sum_{\boldsymbol{k}} g_{\boldsymbol{k}}\; s^i_{\boldsymbol{n}-\boldsymbol{k}} = \sum_{\boldsymbol{m}} s^i_{\boldsymbol{m}}\; g_{\boldsymbol{n}-\boldsymbol{m}}. \end{gathered} \tag{16}$$

This convolution's *frequency response* is the Fourier transform of kernel g,

$$G(\boldsymbol{f}) \triangleq \sum_{\boldsymbol{k}} g_{\boldsymbol{k}}\, \mathrm{e}^{-\mathrm{j}2\pi \boldsymbol{f}\boldsymbol{k}}, \tag{17}$$

and is so named because $s^i \leftrightarrow S^i$ and this relationship $g \leftrightarrow G$ and (16) give

$$\begin{aligned} S^j(\boldsymbol{f}) &= \sum_{\boldsymbol{n}} \Big(\sum_{\boldsymbol{k}} g_{\boldsymbol{k}}\; s^i_{\boldsymbol{n}-\boldsymbol{k}} \Big) \mathrm{e}^{-\mathrm{j}2\pi \boldsymbol{f}\boldsymbol{n}} \\ &= \sum_{\boldsymbol{k}} \sum_{\boldsymbol{n}} g_{\boldsymbol{k}}\, s^i_{\boldsymbol{n}} \mathrm{e}^{-\mathrm{j}2\pi \boldsymbol{f}(\boldsymbol{n}+\boldsymbol{k})} \\ &= S^i(\boldsymbol{f})\; G(\boldsymbol{f}). \end{aligned}$$

Notation h^i, with any identifier i, will always refer to $h$ardware coefficients.

Our second signal operation is *frequency translation*, for which we use

$$s^j = s^i \circlearrowleft \boldsymbol{f}_\Delta \quad \text{for} \quad s^j_{\boldsymbol{n}} = \mathrm{e}^{\mathrm{j}2\pi \boldsymbol{f}_\Delta \boldsymbol{n}}\, s^i_{\boldsymbol{n}}. \tag{18}$$

The arguments to the $\circlearrowleft$ operator are the real two-element row vector $\boldsymbol{f}_\Delta$ on the right, which might be a product, and everything to the left up to an enclosing parenthesis. A sequence of $\star$ and $\circlearrowleft$ operations should be read left to right. The 2D Fourier transform of (18) is just, using (15),

$$S^j(\boldsymbol{f}) = \mathrm{e}^{\mathrm{j}2\pi \boldsymbol{f}_\Delta \boldsymbol{n}} \sum_{\boldsymbol{n}} s^i_{\boldsymbol{n}}\, \mathrm{e}^{-\mathrm{j}2\pi \boldsymbol{f}\boldsymbol{n}} = \sum_{\boldsymbol{n}} s^i_{\boldsymbol{n}}\, \mathrm{e}^{-\mathrm{j}2\pi (\boldsymbol{f}-\boldsymbol{f}_\Delta)\boldsymbol{n}} = S^i(\boldsymbol{f}-\boldsymbol{f}_\Delta). \tag{19}$$

For the related *frequency referral* $\circlearrowright$ of a frequency response, we use

$$g^j = \boldsymbol{f}_\Delta \circlearrowright g^i \quad \text{for} \quad g^j_{\boldsymbol{k}} = g^i_{\boldsymbol{k}}\, \mathrm{e}^{-\mathrm{j}2\pi \boldsymbol{f}_\Delta \boldsymbol{k}}. \tag{20}$$

The arguments of the $\circlearrowleft$ operation are the real two-element row vector $\boldsymbol{f}_\Delta$ to the left, which might be a product, and everything to the right up to an enclosing parenthesis. A sequence of these $\circlearrowleft$ operations should therefore be read right to left. The 2D Fourier transform of (20) is just, using (17),

$$G^j(\boldsymbol{f}) = \mathrm{e}^{-\mathrm{j}2\pi \boldsymbol{f}_\Delta \boldsymbol{k}} \sum_{\boldsymbol{k}} g^i_{\boldsymbol{k}}\, \mathrm{e}^{-\mathrm{j}2\pi \boldsymbol{f} \boldsymbol{k}} = \sum_{\boldsymbol{k}} g^i_{\boldsymbol{k}}\, \mathrm{e}^{-\mathrm{j}2\pi (\boldsymbol{f}+\boldsymbol{f}_\Delta) \boldsymbol{k}} = G^i(\boldsymbol{f}+\boldsymbol{f}_\Delta). \tag{21}$$

Key to our use of the $\circlearrowleft$ operation is *referral identity*

$$s^i \circlearrowleft \boldsymbol{f}_\Delta \star g = s^i \star (\boldsymbol{f}_\Delta \circlearrowleft g) \circlearrowleft \boldsymbol{f}_\Delta, \tag{22}$$

which just expresses, using (16) and (18),

$$\sum_{\boldsymbol{k}} g_{\boldsymbol{k}}\; s^i_{\boldsymbol{n}-\boldsymbol{k}}\; \mathrm{e}^{\mathrm{j}2\pi \boldsymbol{f}_\Delta (\boldsymbol{n}-\boldsymbol{k})} = \mathrm{e}^{\mathrm{j}2\pi \boldsymbol{f}_\Delta \boldsymbol{n}} \sum_{\boldsymbol{k}} \big(g_{\boldsymbol{k}}\; \mathrm{e}^{-\mathrm{j}2\pi \boldsymbol{f}_\Delta \boldsymbol{k}}\big) s^i_{\boldsymbol{n}-\boldsymbol{k}}.$$

From (19) and (21), the Fourier transform of referral identity (22) and its natural interpretation as well is the obvious identity

$$S^i(\boldsymbol{f}-\boldsymbol{f}_\Delta)\, G(\boldsymbol{f}) = \big[S^i(\boldsymbol{f}')\, G(\boldsymbol{f}'+\boldsymbol{f}_\Delta)\big]_{\boldsymbol{f}'=\boldsymbol{f}-\boldsymbol{f}_\Delta}.$$

Our third signal operation is layer-to-layer thinning using 2D decimation operator $\downarrow$. With signal sets and 2×2 integer matrices both arbitrary, we use

$$\begin{aligned} s^j &= s^i \downarrow \mathbf{R} \quad \text{for} \quad s^j_{\boldsymbol{n}} = s^i_{\mathbf{R}\boldsymbol{n}}, \\ s^i \downarrow \mathbf{R}_i \downarrow \mathbf{R}_j &\quad \text{for} \quad ((s^i \downarrow \mathbf{R}_i) \downarrow \mathbf{R}_j). \end{aligned} \tag{23}$$

The arguments to the $\downarrow$ operator are the matrix to the right, which might be a product, and everything to the left up to an enclosing parenthesis. So a sequence of $\star$ and $\circlearrowleft$ and $\downarrow$ operations should be read left to right. By (23),

$$s^i \downarrow \mathbf{R}_i \downarrow \mathbf{R}_j == s^i \downarrow (\mathbf{R}_i \mathbf{R}_j). \tag{24}$$

For the related *zero-interpolation* $\uparrow$ of a frequency response in a three-operand combination with a convolution. we use

$$\begin{aligned} g \;=\; g^i \star \mathbf{R} \uparrow g^j \quad &\text{for} \quad g_{\boldsymbol{n}} = \sum_{\boldsymbol{k}} g^j_{\boldsymbol{k}}\; g^i_{\boldsymbol{n}-\mathbf{R}\boldsymbol{k}}, \\ g^i \star \mathbf{R}_i \uparrow g^j \star \mathbf{R}_j \uparrow g^k \quad &\text{for} \quad (g^i \star \mathbf{R}_i \uparrow (g^j \star \mathbf{R}_j \uparrow g^k)). \end{aligned} \tag{25}$$

The arguments of the $\star\uparrow$ operation are the coefficient set to the left, the matrix in the center, which might be a product, and everything to the right up to an enclosing parenthesis. A sequence of $\star\uparrow$ and $\circlearrowleft$ operations should therefore be read right to left. Key to our use of this operation is a variant on a "Noble identity" routinely used in multi-rate DSP,

$$s \star h \downarrow \mathbf{R} \star g = s \star (h \star \mathbf{R} \uparrow g) \downarrow \mathbf{R}, \tag{26}$$

a shorthand for some simple re-ordering and re-indexing of finite sums:

$$\sum_{\boldsymbol{k}} g_{\boldsymbol{k}} \left[\sum_{m} s_m \; h_{\mathbf{R}\boldsymbol{\ell}-m} \right]_{\boldsymbol{\ell}=\boldsymbol{n}-\boldsymbol{k}} = \qquad \sum_{m} s_m \left[\sum_{\boldsymbol{k}} g_{\boldsymbol{k}} \; h_{\boldsymbol{\ell}-\mathbf{R}\boldsymbol{k}} \right]_{\boldsymbol{\ell}=\mathbf{R}\boldsymbol{n}-m} .$$

The 2D Fourier transform of $g = g^i \star \mathbf{R} \uparrow g^j$ is just, using (17) and (25),

$$\begin{aligned} G(\boldsymbol{f}) &= \sum_{\boldsymbol{n}} \Big(\sum_{\boldsymbol{k}} g^j_{\boldsymbol{k}} \; g^i_{\boldsymbol{n}-\mathbf{R}\boldsymbol{k}} \Big) \mathrm{e}^{-\mathrm{j}2\pi \boldsymbol{f}\boldsymbol{n}} \\ &= \sum_{\boldsymbol{k}} \sum_{\boldsymbol{n}} g^j_{\boldsymbol{k}} \; g^i_{\boldsymbol{n}} \; \mathrm{e}^{-\mathrm{j}2\pi \boldsymbol{f}(\boldsymbol{n}+\mathbf{R}\boldsymbol{k})} \\ &= G^i(\boldsymbol{f}) \; G^j(\boldsymbol{f}\mathbf{R}). \end{aligned} \tag{27}$$

FIR spatial filtering: a beam sum. Let $s^0_{\boldsymbol{n}} \triangleq A(\mathbf{B}\boldsymbol{n}, f)$, the output in (14) of the element at $\mathbf{B}\boldsymbol{n}$, so that

$$s^0_{\boldsymbol{n}} = \int \mathrm{A}(\underline{\mathbf{k}}, f) \; \mathrm{e}^{\mathrm{j}2\pi \underline{\mathbf{x}} \cdot \mathbf{B}\boldsymbol{n}} \, d\underline{\mathbf{k}}, \tag{28}$$

and convolve s^0 with some coefficient set g^1 to obtain

$$s \triangleq s^0 \star g^1. \tag{29}$$

If $g^1 \leftrightarrow G^1$, then (14), (16), and (17) together give

$$\begin{aligned} s_{\boldsymbol{n}}(f) &= \sum_{\boldsymbol{k}} g^1_{\boldsymbol{k}} \, s^0_{\boldsymbol{n}-\boldsymbol{k}} = \sum_{\boldsymbol{k}} g^1_{\boldsymbol{k}} \int \mathrm{A}(\underline{\mathbf{k}}, f) \; \mathrm{e}^{\mathrm{j}2\pi \underline{\mathbf{x}} \cdot \mathbf{B}(\boldsymbol{n}-\boldsymbol{k})} \, d\underline{\mathbf{k}} \\ &= \int \mathrm{A}(\underline{\mathbf{k}}, f) \; G^1(\underline{\mathbf{k}} \cdot \mathbf{B}) \; \mathrm{e}^{\mathrm{j}2\pi \underline{\mathbf{x}} \cdot \mathbf{B}\boldsymbol{n}} \, d\underline{\mathbf{k}}. \end{aligned} \tag{30}$$

Comparing with (28) shows that $s_{\boldsymbol{n}}(f)$ is just $s^0_{\boldsymbol{n}}(f)$ with but with directional ($\underline{\mathbf{k}}$ sensitive) selectivity applied through $G^1(\underline{\mathbf{k}} \cdot \mathbf{B})$. Here we are only interested in one signal from set s, the *array output* s_0, given by (30) to be

$$s_0(f) = \int \mathrm{A}(\underline{\mathbf{k}}, f) \; G^1(\underline{\mathbf{k}} \cdot \mathbf{B}) \; d\underline{\mathbf{k}}. \tag{31}$$

Frequency response $G^1(\boldsymbol{f})$ from (17) is that of a 2D FIR (finite impulse response) filter. Spatial frequency response $G^1(\underline{\mathbf{k}} \cdot \mathbf{B})$ is an *array factor*. This name is conventional, give or take an irrelevant index sign, as convolution $s_{\boldsymbol{n}} = \sum_{\boldsymbol{k}} g^1_{\boldsymbol{k}} \star s^0_{\boldsymbol{n}-\boldsymbol{k}}$ specializes when $\boldsymbol{n} = 0$ to ordinary linear beam sum

$$s_0(f) = \sum_{\boldsymbol{k}} g^1_{\boldsymbol{k}} \, s^0_{-\boldsymbol{k}}(f). \tag{32}$$

We use the term array factor more generally for any dimensionless scalar spatial frequency response determined by one or more coefficient sets.

Coefficient set g^1 is finite, so beam sum (32) depends on only a finite number of element outputs $s^0_{-\boldsymbol{k}}(f)$, and as a practical matter the periodic structure of the physical array is truncated at a point where doing so does not materially affect the fields integrated to get that finite subset of element outputs. The array remains infinite in the analysis for simplicity.

Steering the array factor. Changing the array's directional selectivity in (31) by changing coefficients g^1 in (29) is sometimes a reasonable option. A standard alternative is to keep coefficients g^1 fixed but replace (29) with

$$s \triangleq s^0 \circlearrowleft \underline{\mathbf{k}}_\Delta \cdot \mathbf{B} \star g^1, \tag{33}$$

where $\underline{\mathbf{k}}_\Delta$ can be easily varied. By referral identity (22) this is equivalent to

$$s = s^0 \star (\underline{\mathbf{k}}_\Delta \cdot \mathbf{B} \circlearrowright g^1) \circlearrowleft \underline{\mathbf{k}}_\Delta \cdot \mathbf{B}$$

By (18) the $\boldsymbol{n} = 0$ signal is unaffected by the trailing $\circlearrowleft \underline{\mathbf{k}}_\Delta \cdot \mathbf{B}$, so array output s_0 is exactly as in (31) but with Fourier pair $g^1 \leftrightarrow G^1(\boldsymbol{f})$ replaced with, using (21), Fourier pair $\underline{\mathbf{k}}_\Delta \cdot \mathbf{B} \circlearrowright g^1 \leftrightarrow G^1(\boldsymbol{f} + \underline{\mathbf{k}}_\Delta \cdot \mathbf{B})$. It follows that

$$s_0(f) = \int \mathrm{A}(\underline{\mathbf{k}}, f)\, G^1((\underline{\mathbf{k}} + \underline{\mathbf{k}}_\Delta) \cdot \mathbf{B})\ d\underline{\mathbf{k}}. \tag{34}$$

This is just (31) but with directional selectivity now applied through translated or *steered* array factor $G^1((\underline{\mathbf{k}} + \underline{\mathbf{k}}_\Delta) \cdot \mathbf{B})$.

By what $\underline{\mathbf{k}}_\Delta$ should the array factor be steered? Recall that on the Helmholz cone $\mathbf{k} = \widehat{k} f/c$, where unit vector $\widehat{k}$ is the look direction. Typically FIR frequency response $G^1(\boldsymbol{f})$ is designed to be lowpass in character so that unsteered array factor $G^1(\underline{\mathbf{k}} \cdot \mathbf{B}) = G^1(\widehat{k} \cdot \mathbf{B} f/c)$ has the highest gain for directions near boresight direction $\widehat{b}$, so to have it instead maximize reception from look direction $\widehat{k}_0$, we need only set $\underline{\mathbf{k}}_\Delta = -\widehat{k}_0 f/c$. The steered array factor is then $G^1((\underline{\mathbf{k}} + \underline{\mathbf{k}}_\Delta) \cdot \mathbf{B}) = G^1((\widehat{k} - \widehat{k}_0) \cdot \mathbf{B} f/c)$.

Hardware realization of $s_0(f)$ is per (32) except that, comparing (28) to (33) and using frequency-translation definition (18), element output $s^0_{\boldsymbol{n}}(f)$ must be replaced with *steered element output* $\mathrm{e}^{\mathrm{j}2\pi \underline{\mathbf{k}}_\Delta \cdot \mathbf{B}\boldsymbol{n}}\, s^0_{\boldsymbol{n}}(f)$, resulting in

$$s_0(f) = \sum_{\boldsymbol{k}} g^1_{\boldsymbol{k}}\, \mathrm{e}^{\mathrm{j}2\pi \underline{\mathbf{k}}_\Delta \cdot \mathbf{B}\boldsymbol{n}}\, s^0_{\boldsymbol{n}}(f). \tag{35}$$

Given $\underline{\mathbf{k}}_\Delta = -\widehat{k}_0 f/c$, element output $\boldsymbol{n}$ has either been delayed by $\widehat{k}_0 \cdot \mathbf{B}\boldsymbol{n}/c$ or, at some narrowband operating frequency $f = f_0$, had its phase shifted by $-2\pi \widehat{k}_0 \cdot \mathbf{B}\boldsymbol{n}/c \bmod 2\pi$. (It is to be always understood that additional delay, constant across $\boldsymbol{n}$, is generally added to make each net operation causal.)

Example: a seven-element triangular array. The simple array schematic in Fig. 1 establishes our diagramming conventions. On the left, boresight direction $\widehat{b}$ is downward, and seven elements are represented as gray pancakes centered on lattice points in the horizontal array plane at the bottom. Only elements included in the beam sum are shown (or counted), but in a physical array the periodic array structure would be continued for some modest distance using terminated "guard elements." In the diagram each element output flows up through a barrel-shaped symbol representing both a low-noise amplifier (LNA)

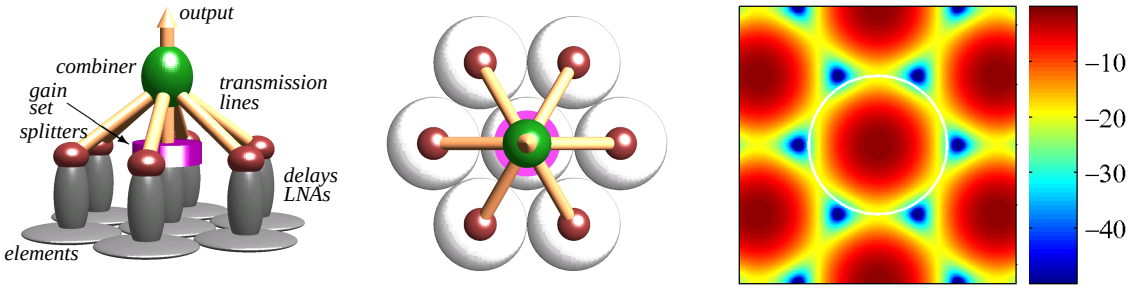

Figure 1: A 3D block diagram of a seven-element triangular receive array viewed from behind the array plane and the dB magnitude of its unsteered array factor versus 2D spatial frequency $\underline{\mathbf{k}}$ at $f = f_u$. The visible region is inside the white circle. Steering this array will translate some of the six stopbands (sidelobe regions), here shown just outside the circle, into the visible region.

and a delay or phase shift for steering. These feed transmission lines, mostly via power splitters, functionally irrelevant here but useful later, shown as prolate spheroids. A combiner, shown as an oblate spheroid, realizes the array output as a coefficient-weighted beam sum. The weighting means are realization specific and not always shown, though here the center combiner input is marked by a toroidal gain set as using a different weight than do the six outer combiner inputs. Inputs given identical weights by symmetric coefficients may of course actually be combined before weights are applied. With triangular array structures very efficient six- and twelve-way coefficient symmetries are available.

The 3D feed system of elements, splitters, combiners, and transmission lines shown is an abstract representation . We do not exclude any technology that might help realize the operations desired. Planar structures might be used for elements and transmission lines, for example, or space-fed discrete lens arrays [5] might realize entire arrays or subarray layers.

2.4. Conventional Subarrayed Systems

Sublattices. Suppose a lattice $\mathbf{B}'\mathbb{Z}^2$ is a subset of lattice $\mathbf{B}\mathbb{Z}^2$ and therefore a *sublattice* of $\mathbf{B}\mathbb{Z}^2$. Basis vectors $\mathbf{B}'\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ and $\mathbf{B}'\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ of sublattice $\mathbf{B}'\mathbb{Z}^2$ must be in lattice $\mathbf{B}\mathbb{Z}^2$ and therefore representable as $\mathbf{B}'\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right] = \mathbf{B}\boldsymbol{z}_1$ and $\mathbf{B}'\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right] = \mathbf{B}\boldsymbol{z}_2$ for some vectors $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathbb{Z}^2$. It follows that $\mathbf{B}' = \mathbf{B}'\left[\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right] = \mathbf{B}[\boldsymbol{z}_1, \boldsymbol{z}_2]$, and we term integer matrix $[\boldsymbol{z}_1, \boldsymbol{z}_2]$ the *resampling matrix* relating generating matrices $\mathbf{B}$ and $\mathbf{B}'$. Any nonsingular 2×2 integer matrix $\mathbf{R}$ can serve as a resampling matrix to resample a lattice $\mathbf{B}\mathbb{Z}^2$ to create a sublattice $\mathbf{BR}\mathbb{Z}^2$. Lattice bases are not unique, so resampling matrices are not unique either.

If $\mathbf{G} = \mathbf{B}^T \cdot \mathbf{B}$ is the Gram matrix of lattice basis $\mathbf{B}$, then the Gram matrix of sublattice basis $\mathbf{BR}$ is $\mathbf{R}^T\mathbf{B}^T \cdot \mathbf{BR} = \mathbf{R}^T\mathbf{GR}$. If the two Gram matrices are identical to within a scale factor, then sublattice $\mathbf{BR}\mathbb{Z}^2$ is necessarily *similar* to lattice $\mathbf{B}\mathbb{Z}^2$. If the period of lattice $\mathbf{B}\mathbb{Z}^2$ has area $|\mathbf{G}|^{1/2}$, then the period of sublattice $\mathbf{BR}\mathbb{Z}^2$ has area $|\mathbf{R}^T\mathbf{GR}|^{1/2} = |\mathbf{G}|^{1/2}|\mathbf{R}|$. Lattice $\mathbf{B}\mathbb{Z}^2$ is therefore integer $|\mathbf{R}|$ times as dense as sublattice $\mathbf{BR}\mathbb{Z}^2$.

The resampling matrices used in this paper are easily constructed. Given lattice basis $\mathbf{B} \triangleq [\mathbf{b}_1, \mathbf{b}_2]$ there is a 2×2 integer matrix $\mathbf{S}$, shown in Fig. 2 for $\mathbf{B} = \mathbf{B}_\square$ and $\mathbf{B} = \mathbf{B}_\triangle$, with $|S| = 1$ so that map $\boldsymbol{z} \mapsto \mathbf{S}\boldsymbol{z}$ is a rotation defined on $\mathbb{Z}^2$ and map $\mathbf{B}\boldsymbol{z} \mapsto \mathbf{BS}\boldsymbol{z}$ is a rotation defined on $\mathbf{B}\mathbb{Z}^2$, that rotates $\mathbf{b}_1$ to $\mathbf{b}_2$. The latter is just $\mathbf{BS}\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right] = \mathbf{B}\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$, which is equivalent

$\mathbf{B}$	$\mathbf{S}$	1	2	3	4
$\mathbf{B}_\square$	$\left[\begin{smallmatrix}0&-1\\1&0\end{smallmatrix}\right]$	$\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$	$\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$		$\left[\begin{smallmatrix}2\\0\end{smallmatrix}\right]$
$\mathbf{B}_\triangle$	$\left[\begin{smallmatrix}0&-1\\1&1\end{smallmatrix}\right]$	$\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$		$\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$	$\left[\begin{smallmatrix}2\\0\end{smallmatrix}\right]$

5	6	7	8	9	10	11
$\left[\begin{smallmatrix}2\\1\end{smallmatrix}\right]$			$\left[\begin{smallmatrix}2\\2\end{smallmatrix}\right]$	$\left[\begin{smallmatrix}3\\0\end{smallmatrix}\right]$	$\left[\begin{smallmatrix}3\\1\end{smallmatrix}\right]$	
		$\left[\begin{smallmatrix}2\\1\end{smallmatrix}\right]$		$\left[\begin{smallmatrix}3\\0\end{smallmatrix}\right]$		

12	13	$\cdots$	16	17	18	19
	$\left[\begin{smallmatrix}3\\2\end{smallmatrix}\right]$		$\left[\begin{smallmatrix}4\\0\end{smallmatrix}\right]$	$\left[\begin{smallmatrix}4\\1\end{smallmatrix}\right]$	$\left[\begin{smallmatrix}3\\3\end{smallmatrix}\right]$	
$\left[\begin{smallmatrix}2\\2\end{smallmatrix}\right]$	$\left[\begin{smallmatrix}3\\1\end{smallmatrix}\right]$		$\left[\begin{smallmatrix}4\\0\end{smallmatrix}\right]$			$\left[\begin{smallmatrix}3\\2\end{smallmatrix}\right]$

Figure 2: For each basis $\mathbf{B}$, matrix $\mathbf{S}$ creates resampling matrix $\mathbf{R}=\left[\boldsymbol{z},\mathbf{S}\boldsymbol{z}\right]$ from the vector $\boldsymbol{z}\in\mathbb{Z}^2$ tabulated on the right under the desired value of $|\mathbf{R}|$.

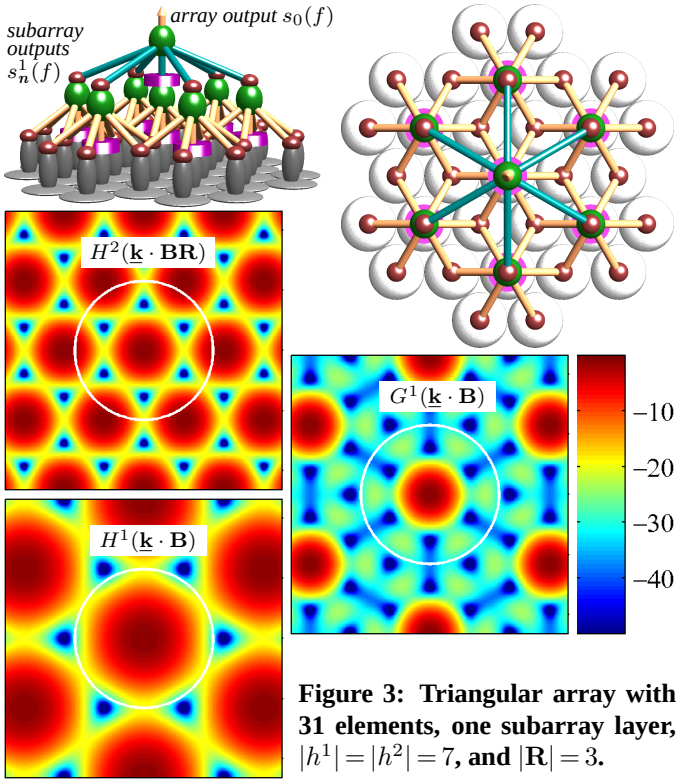

Figure 3: Triangular array with 31 elements, one subarray layer, $|h^1|=|h^2|=7$, and $|\mathbf{R}|=3$.

to $\mathbf{S}\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]=\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ by the linear independence of $\mathbf{b}_1$ and $\mathbf{b}_2$. If some specific but arbitrary lattice point $\mathbf{B}\boldsymbol{z}$ and rotated point $\mathbf{BS}\boldsymbol{z}$ are used as the two sublattice basis vectors, the resulting sublattice basis $\mathbf{BR}$ has $\mathbf{R}=\left[\boldsymbol{z},\mathbf{S}\boldsymbol{z}\right]$ and yields a sublattice $\mathbf{BR}\mathbb{Z}^2$ similar to $\mathbf{B}\mathbb{Z}^2$ with the area of the period scaled up by $\|\mathbf{B}\boldsymbol{z}\|^2=\boldsymbol{z}^T\mathbf{G}\boldsymbol{z}=|\mathbf{R}|$. For bases $\mathbf{B}=\mathbf{B}_\square$ and $\mathbf{B}=\mathbf{B}_\triangle$, Fig. 2 lists the vectors $\boldsymbol{z}$ we use to obtain all possible density ratios $|\mathbf{R}|$ up to 19.

Use subarrays: translate, filter, and decimate. Replace (33) with

$$s^1 \triangleq s^0 \circlearrowleft \underline{\mathbf{k}}_1\cdot\mathbf{B} \star h^1 \downarrow \mathbf{R} \tag{36}$$

$$s \triangleq s^1 \circlearrowleft \underline{\mathbf{k}}_2\cdot\mathbf{BR} \star h^2. \tag{37}$$

Realization of *subarray outputs* s^1 is traditionally analog, and digital combining typically then realizes array output $s_0(f)$, given by (37) to be

$$s_0(f) = \sum_{\boldsymbol{k}} h^2_{\boldsymbol{k}}\, s^1_{-\boldsymbol{k}}(f). \tag{38}$$

Combining (36) and (37) and then applying, right to left, referral identity (22), Noble identity (26), and then referral identity (22) again yields

$$\begin{aligned} s &= s^0 \circlearrowleft \underline{\mathbf{k}}_1\cdot\mathbf{B} \star h^1 \downarrow \mathbf{R} \star (\underline{\mathbf{k}}_2\cdot\mathbf{BR} \circlearrowleft h^2) \circlearrowleft \underline{\mathbf{k}}_2\cdot\mathbf{BR} \\ &= s^0 \circlearrowleft \underline{\mathbf{k}}_1\cdot\mathbf{B} \star (h^1 \star \mathbf{R} \uparrow \underline{\mathbf{k}}_2\cdot\mathbf{BR} \circlearrowleft h^2) \downarrow \mathbf{R} \circlearrowleft \underline{\mathbf{k}}_2\cdot\mathbf{BR} \\ &= s^0 \star (\underline{\mathbf{k}}_1\cdot\mathbf{B} \circlearrowleft h^1 \star \mathbf{R} \uparrow \underline{\mathbf{k}}_2\cdot\mathbf{BR} \circlearrowleft h^2) \circlearrowleft \underline{\mathbf{k}}_1\cdot\mathbf{B} \downarrow \mathbf{R} \circlearrowleft \underline{\mathbf{k}}_2\cdot\mathbf{BR}. \end{aligned} \tag{39}$$

Trailing frequency translation and decimation in (39) do not affect the $\boldsymbol{n} = 0$ variable, so (32) and (31) hold but with g^1 the parenthesized quantity above,

$$g^1 = \underline{\mathbf{k}}_1 \cdot \mathbf{BR} \circlearrowleft h^1 \star \mathbf{R} \uparrow \underline{\mathbf{k}}_2 \cdot \mathbf{BR} \circlearrowleft h^2.$$

To derive its Fourier transform, apply, right to left on larger and larger subexpressions, first (21), then (27), and finally (21) again to obtain first $H^2(\boldsymbol{f})$, then $H^2(\boldsymbol{f} + \underline{\mathbf{k}}_2 \cdot \mathbf{BR})$, then $H^1(\boldsymbol{f})\, H^2(\boldsymbol{f}\mathbf{R} + \underline{\mathbf{k}}_2 \cdot \mathbf{BR})$, and finally $G^1(\boldsymbol{f}) = H^1(\boldsymbol{f} + \underline{\mathbf{k}}_1 \cdot \mathbf{B})\, H^2((\boldsymbol{f} + \underline{\mathbf{k}}_1 \cdot \mathbf{B})\mathbf{R} + \underline{\mathbf{k}}_2 \cdot \mathbf{BR})$. Substituting *overall array factor* $G^1(\underline{\mathbf{k}} \cdot \mathbf{B})$ and $\mathrm{A}(\underline{\mathbf{k}}, f)$ of (13) into (31) gives array output

$$\begin{aligned} s_0(f) = \int \Bigg[\int \overrightarrow{\mathbf{E}}\Big(\big[\underline{\mathbf{k}} + k\widehat{b}, f/c \big] \Big)\, dk \Bigg] \\ \cdot \Big(\quad \overrightarrow{\underline{\mathbf{\Phi}}}(\underline{\mathbf{k}}, f) \\ \times H^1((\underline{\mathbf{k}} + \underline{\mathbf{k}}_1 \quad) \cdot \mathbf{B} \quad) \\ \times H^2((\underline{\mathbf{k}} + \underline{\mathbf{k}}_1 + \underline{\mathbf{k}}_2) \cdot \mathbf{BR}) \Big)\ d\underline{\mathbf{k}}. \end{aligned} \tag{40}$$

The inner integral is the incident array-plane electric field, and directional selectivity is applied to it through the dot product with the three functions of $\underline{\mathbf{k}}$ in the large parentheses. Products of those functions have standard names.

(periodic) element pattern	$\overrightarrow{\underline{\mathbf{\Phi}}}(\underline{\mathbf{k}}, f)$
subarray pattern	$\overrightarrow{\underline{\mathbf{\Phi}}}(\underline{\mathbf{k}}, f)\, H^1((\underline{\mathbf{k}} + \underline{\mathbf{k}}_1) \cdot \mathbf{B})$
overall array pattern	$\overrightarrow{\underline{\mathbf{\Phi}}}(\underline{\mathbf{k}}, f)\, H^1((\underline{\mathbf{k}} + \underline{\mathbf{k}}_1) \cdot \mathbf{B})\, H^2((\underline{\mathbf{k}} + \underline{\mathbf{k}}_1 + \underline{\mathbf{k}}_2) \cdot \mathbf{BR})$
subarray array factor	$H^1((\underline{\mathbf{k}} + \underline{\mathbf{k}}_1) \cdot \mathbf{B})$
digital array factor	$H^2((\underline{\mathbf{k}} + \underline{\mathbf{k}}_1 + \underline{\mathbf{k}}_2) \cdot \mathbf{BR})$
overall array factor	$H^1((\underline{\mathbf{k}} + \underline{\mathbf{k}}_1) \cdot \mathbf{B})\, H^2((\underline{\mathbf{k}} + \underline{\mathbf{k}}_1 + \underline{\mathbf{k}}_2) \cdot \mathbf{BR})$

From here on we fix and ignore $\overrightarrow{\underline{\mathbf{\Phi}}}(\underline{\mathbf{k}}, f)$. We design the overall array factor.

Location of subarray hardware. We nominally locate signal $s^0_{\boldsymbol{n}}$ at the location $\mathbf{B}\boldsymbol{n}$ of the element from which it is taken. Then in (36)

$$s^1_{\boldsymbol{n}} = \sum_{\boldsymbol{k}} h^1_{\boldsymbol{k}}\, \mathrm{e}^{\mathrm{j}2\pi(\underline{\mathbf{k}}_1 \cdot \mathbf{B})(\mathbf{R}\boldsymbol{n} - \boldsymbol{k})}\, s^0_{\mathbf{R}\boldsymbol{n} - \boldsymbol{k}} \tag{41}$$

depends on input $s^0_{\mathbf{R}\boldsymbol{n} - \boldsymbol{k}}$ located at $\mathbf{BR}\boldsymbol{n} - \mathbf{B}\boldsymbol{k}$ if $h^1_{\boldsymbol{k}} \neq 0$. Coefficient set h^1 is finite, so there is some radius $r > 0$ such that $\|\mathbf{B}\boldsymbol{k}\| \leq r$ for every $\boldsymbol{k}$ with $h^1_{\boldsymbol{k}} \neq 0$. Elements contributing to $s^1_{\boldsymbol{n}}$ are inside an array-plane circle with (often small) radius r centered at $\mathbf{BR}\boldsymbol{n}$, so we assign this nominal location to $s^1_{\boldsymbol{n}}$ because the hardware to realize it would generally be placed near this point for efficiency. Similarly, (38) locates array output $s_0(f)$ at the origin.

Realizing the directional selectivity of (40) using a digital beam sum (38) of subarray outputs s^1, instead of direct DSP computation of s_0 in (32), is often done specifically to lower the array-plane density of signals that must be digitized by a factor of $|\mathbf{R}|$.

To assign locations to the coefficients, rewrite (41) with the signals of s^1 and s^0 not indexed but functionally dependent on their locations:

$$s^1(\mathbf{B}\mathbf{R}\boldsymbol{n}) = \sum_{\boldsymbol{k}} h^1_{\boldsymbol{k}}\, \mathrm{e}^{\mathrm{j}2\pi \underline{\mathbf{k}}_1\cdot(\mathbf{B}\mathbf{R}\boldsymbol{n}-\mathbf{B}\boldsymbol{k})}\, s^0(\mathbf{B}\mathbf{R}\boldsymbol{n}-\mathbf{B}\boldsymbol{k}). \tag{42}$$

Replacing $h^1_{\boldsymbol{k}}$ with $h^1(\mathbf{B}\boldsymbol{k})$ makes this a spatial convolution, so we assign location $\mathbf{B}\boldsymbol{k}$ to $h^1_{\boldsymbol{k}}$. A similar argument using (38) assigns location $\mathbf{B}\mathbf{R}\boldsymbol{k}$ to $h^2_{\boldsymbol{k}}$. We'll formally convolve in index space only, but it is sometimes intuitively valuable to think of convolutions spatially.

What each array factor contributes to a simple example system. In Fig. 3 subarray-output set $s^1_{\boldsymbol{n}}$ has seven signals, and each is a combination, coincidentally, of seven element outputs. We say there are seven overlapping seven-element subarrays. We address the structure of array factors more formally in Section 3, but this simple example design reveals the general idea. Digital array factor $H^2(\underline{\mathbf{k}}\cdot\mathbf{B}\mathbf{R})$ has one passband per period. Coefficients h^1 and h^2 are located on lattice $\mathbf{B}\mathbb{Z}^2$ and its sublattice $\mathbf{B}\mathbf{R}\mathbb{Z}^2$ respectively, and the former is $|\mathbf{R}|$ times more dense, so a period of subarray array factor $H^1(\underline{\mathbf{k}}\cdot\mathbf{B})$ has $|\mathbf{R}| = 3$ times the area of a period of digital array factor $H^2(\underline{\mathbf{k}}\cdot\mathbf{B}\mathbf{R})$. Each period of subarray array factor $H^1(\underline{\mathbf{k}}\cdot\mathbf{B})$ therefore contains three passbands of digital array factor $H^2(\underline{\mathbf{k}}\cdot\mathbf{B}\mathbf{R})$ and is designed to pass one of them but to suppress the other two with small, deep stopbands.

The design of digital array factor $H^2(\underline{\mathbf{k}}\cdot\mathbf{B}\mathbf{R})$ is always about passband or main-beam shape, because it functions as a *shaping filter*, determining the shape of the main beam in the overall array factor. Its small period makes a small main beam relatively easy to achieve. The design of subarray array factor $H^1(\underline{\mathbf{k}}\cdot\mathbf{B})$ is always about the location, size, and depth of its stopbands, because it functions as a *masking filter*, masking periodically repeated digital-array-factor passbands in the overall array factor $G^1(\underline{\mathbf{k}}\cdot\mathbf{B})$. Its large period, relative to stopband size, makes adequate stopband suppression possible with small $|h^1|$, an advantage given the high hardware cost associated with the high spatial density of h^1.

Subarray overlap. The *overlap* is the average number of subarray outputs to which an element away from the edge of the array contributes. In Fig. 3 a third of such elements contribute to a single subarray output each, and $\frac{2}{3}$ contribute to three subarray outputs each, for a $\frac{1}{3} + 3 \times \frac{2}{3} = \frac{7}{3}$ overlap. This can be computed more easily as $|h^1|/|\mathbf{R}|$, which estimates per-band design control because $|h^1|$ subarray coefficients control the behavior of one passband and $|\mathbf{R}|-1$ stopbands. In the design of subarray array factors then, overlap is a figure of merit. The square root (here ≈ 1.53) of the overlap, the overlap *per dimension* or *in each direction*, is often given instead.

Steering the digital array factor separately. In (40) both array factors are steered by *subarray steering vector* $\underline{\mathbf{k}}_1$, which therefore is the primary means of steering the array.

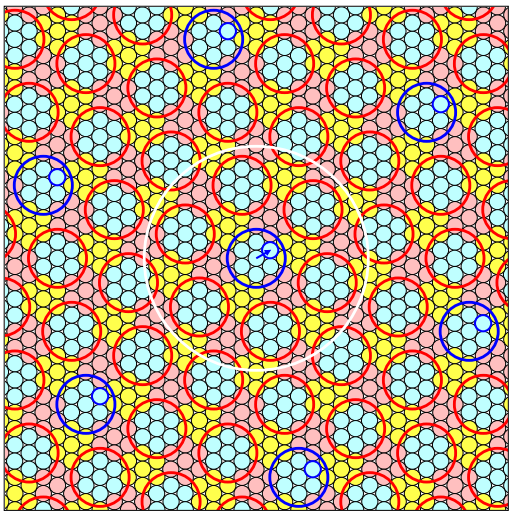

Figure 4: Multiple digital steering vectors $\underline{\mathbf{k}}_2$ (one shown) create a cluster (here 7) of digital beams for each subarray steering vector $\underline{\mathbf{k}}_1$ (here zero, not shown), requiring larger stopbands in the subarray array factor.

Only the digital array factor is steered by *digital steering vector* $\underline{\mathbf{k}}_2$. This leads to two distinct types of system designs.

Single-beam systems set $\underline{\mathbf{k}}_2 = 0$ and don't implement the $\circlearrowleft \underline{\mathbf{k}}_2$ processing step in (37) at all. As in the Fig. 3 example, these systems have stopbands in the subarray array factor just large enough to suppress the correspondingly located passbands of the digital array factor. Subarray steering vector $\underline{\mathbf{k}}_1$ is used to steer the single beam.

Beam-cluster systems realize (37) in parallel for each beam in a cluster, simultaneously producing several outputs s_0 using different fixed values of digital steering vector $\underline{\mathbf{k}}_2$ but the same coefficient set h^2. The stopbands in the subarray array factor must be large enough to suppress the entire cluster of digitally steered beams in the digital array factor. Subarray steering vector $\underline{\mathbf{k}}_1$ steers the entire cluster as a unit.

An example schema for a beam-cluster system is shown in Fig. 4. Here $\mathbf{B}$ is a rotated version (for ease of drawing) of $\mathbf{B}_\triangle$, and resampling uses $\mathbf{R} = 3\mathbf{I}$. (How these relate to passband and stopband locations will be discussed in Section 3.1.) Each of the $|\mathbf{R}| - 1 = 8$ stopbands in each period of the subarray array factor must be large enough to suppress all seven digital beams in the cluster. The stress on the design of the subarray array factor caused by larger passband and stopband sizes requires greater subarray overlap.

The poor man's subarray system. A *poor man's subarray system* uses a single subarray layer with $\mathbf{B} = \mathbf{B}_\square$ and $\mathbf{R} = R\mathbf{I}$ for some $R \in \mathbb{Z}$, has some integer γ overlap in each direction, and uses separable subarray coefficients $h^1_{\boldsymbol{n}} = w_x(n_1)\, w_y(n_2)$, where $\boldsymbol{n} = \left[\begin{smallmatrix} n_1 \\ n_2 \end{smallmatrix}\right]$. Such systems are properly named for those lacking resources because such a system's subarray coefficients h^1 can be constructed by hand or using simple 1D FIR filter-design programs like matlab's `remez` function and because these systems' relatively high cost per unit performance threatens to leave taxpayers in the poorhouse if these architectures are realized

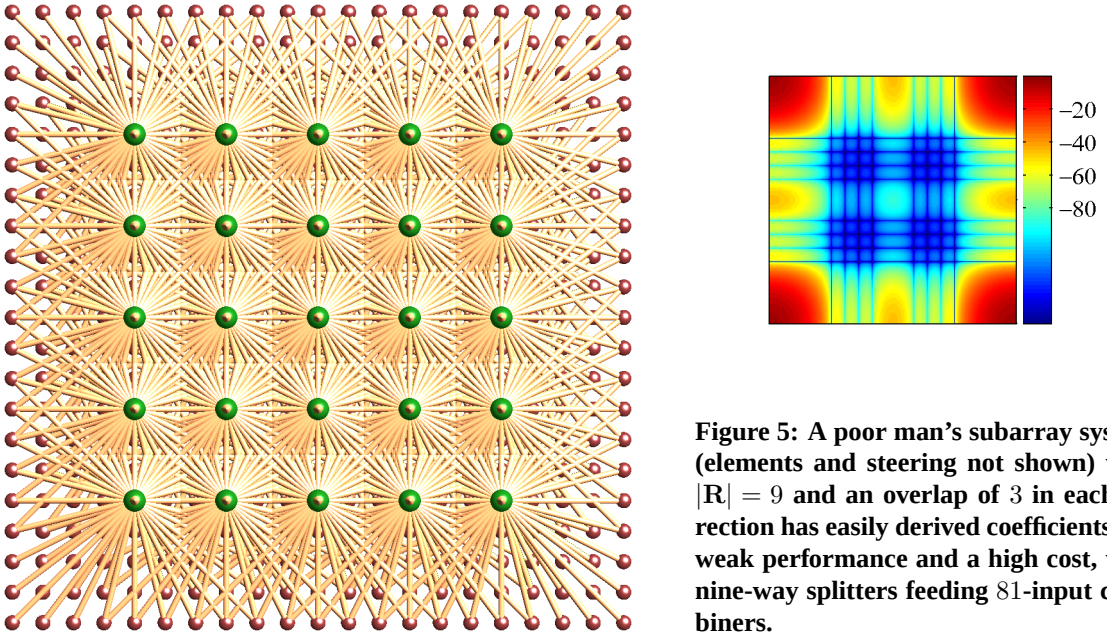

Figure 5: A poor man's subarray system (elements and steering not shown) with $|\mathbf{R}| = 9$ and an overlap of 3 in each direction has easily derived coefficients but weak performance and a high cost, with nine-way splitters feeding 81-input combiners.

for large beam-cluster defense radar systems. In the radar community today the term "subarray" evokes just these unsophisticated schemes, but their proper use in large arrays is to obtain performance baselines, because performing poorly is what they are good at.

One suitably poor way to obtain subarray coefficients h^1 is to begin with

$$g(n) \triangleq \begin{cases} \frac{1}{R} & \text{for } n \in \{0, \ldots, R-1\} \\ 0 & \text{otherwise,} \end{cases}$$

which has Fourier transform

$$G(f) = \frac{1}{R} \sum_{n=0}^{R-1} \mathrm{e}^{-\mathrm{j}2\pi f n} = \frac{1}{R} \frac{1 - \mathrm{e}^{-\mathrm{j}2\pi f R}}{1 - \mathrm{e}^{-\mathrm{j}2\pi f}} = \mathrm{e}^{-\mathrm{j}\pi f (R-1)} \frac{\sin(\pi R f)}{R \sin(\pi f)}.$$

The linear-phase factor $\mathrm{e}^{-\mathrm{j}\pi f(R-1)}$ is easily removed when R is odd by replacing $g(n)$ below with $g\big(n - \frac{R-1}{2}\big)$ to center it about the origin. An even R is trickier. It is simplest to leave the factor there but to ignore it.

To proceed toward coefficients, choose a set $\mathcal{F}$ of a small number $|\mathcal{F}|$ of closely spaced frequencies, each much smaller than $1/R$, to form product

$$G'(f) \triangleq \prod_{\Delta f \in \mathcal{F}} G(f - \Delta f).$$

These frequencies can be "tuned" later for stopband width, depth, and flatness. Function $G'(f)$ is the Fourier transform of the convolution

$$g'(n) = \Big(\bigstar_{\Delta f \in \mathcal{F}}\Big)\, g(n)\, \mathrm{e}^{\mathrm{j}2\pi \Delta f n}.$$

For our poor coefficients we use just $h^1_{\boldsymbol{n}} = g'(n_1)\, g'(n_2)$. The support of $g'(n)$ is $0 \ldots |\mathcal{F}|(R–1)$, so overlap $\gamma = |h^1|/R^2 = \big(1 + |\mathcal{F}|\,(R{-}1)\big)^2/R^2 = \big(|\mathcal{F}| - (|\mathcal{F}|{-}1)/R\big)^2$, an integer in each direction only if R divides $|\mathcal{F}|-1$.

In the Fig. 5 example, $R=3$ and $\mathcal{F} = \frac{0.5}{R(3-(-3))}\{-3,-1,1,3\}$, so each splitter feeds $\gamma = \big(4-(4{-}1)/3\big)^2 = 3^2$ inputs, of which each combiner has $|h^1| = \gamma|\mathbf{R}| = 81$. One array-factor period is shown (with a color scale unique to this plot). The four center stopbands are deeply suppressed, but the four stopbands on the edges are suppressed only half as far in dB terms. The passband, in the corners, varies greatly. The coefficients are poor.

The lattice formulation of subarray systems discussed thus far and its extension to multiple subarray layers, presented next, was motivated by the poor cost performance of these standard, "poor" approaches.

3. Multi-layer Subarrays

The next section develops a multi-layer subarray system as a chain of interleaved spatial filtering and spatial decimation steps. Section 3.2 then addresses array-factor design for individual layers.

3.1. Filter-Decimate Chains

A formal mathematical description. Let *ultimate output* s_0 and the output s^i of subarray layer $i \in \{1, \ldots, N\}$ be as per recursions

$$s^i \triangleq s^{i-1} \circlearrowleft \underline{\mathbf{k}}_i \quad \cdot \mathbf{B}_i \quad \star\, h^i \downarrow \mathbf{R}_i. \tag{43}$$

$$s \triangleq s^N \;\circlearrowleft \underline{\mathbf{k}}_{N+1} \cdot \mathbf{B}_{N+1} \star\, h^{N+1} \tag{44}$$

$$\mathbf{B}_i \triangleq \mathbf{B}_{i-1}\, \mathbf{R}_{i-1} \tag{45}$$

initiated with $s^0_{\boldsymbol{n}} \triangleq A(\mathbf{B}\boldsymbol{n}, f)$ and $\mathbf{B}_1 = \mathbf{B}$, from which $\mathbf{B}_i = \mathbf{B}\mathbf{R}_1 \cdots \mathbf{R}_{i-1}$ is immediate. Subarray output sets s^N and s^{N-1} are the *penultimate* and *ante-penultimate* outputs respectively, and coefficient sets h^{N+1}, h^N, and h^{N-1} are the *ultimate*, *penultimate*, and *ante-penultimate* coefficients respectively.

The corresponding hardware structure. Typically only $\underline{\mathbf{k}}_1$ and perhaps one particular $\underline{\mathbf{k}}_j$ would be nonzero among steering vectors $\underline{\mathbf{k}}_1 \ldots \underline{\mathbf{k}}_{N+1}$, with $\underline{\mathbf{k}}_1$ steering the entire array and with $\underline{\mathbf{k}}_j$, if nonzero, used to form a beam cluster. Any stage could be either analog or digital, but typically j would index the first stage of all-digital processing so that signal sets $s_0 \ldots s_{j-1}$ would be realized in analog hardware, signal set s_{j-1} would be converted from analog to digital, and signal sets $s_j \ldots s_{N+1}$ would be realized digitally. In that case the processing steps in $\circlearrowleft \underline{\mathbf{k}}_i \cdot \mathbf{B}_i \star h^i \downarrow \mathbf{R}_i$ would operate on analog signals for $i = 1 \ldots j{-}1$ and on digital signals for $i = j \ldots N{+}1$.

In Fig. 6 most of the hardware realization for a specific $N=2$ example is sketched. Only $\underline{\mathbf{k}}_1$ steering delays are included, and only the subarray structure is shown. Beginning with

*

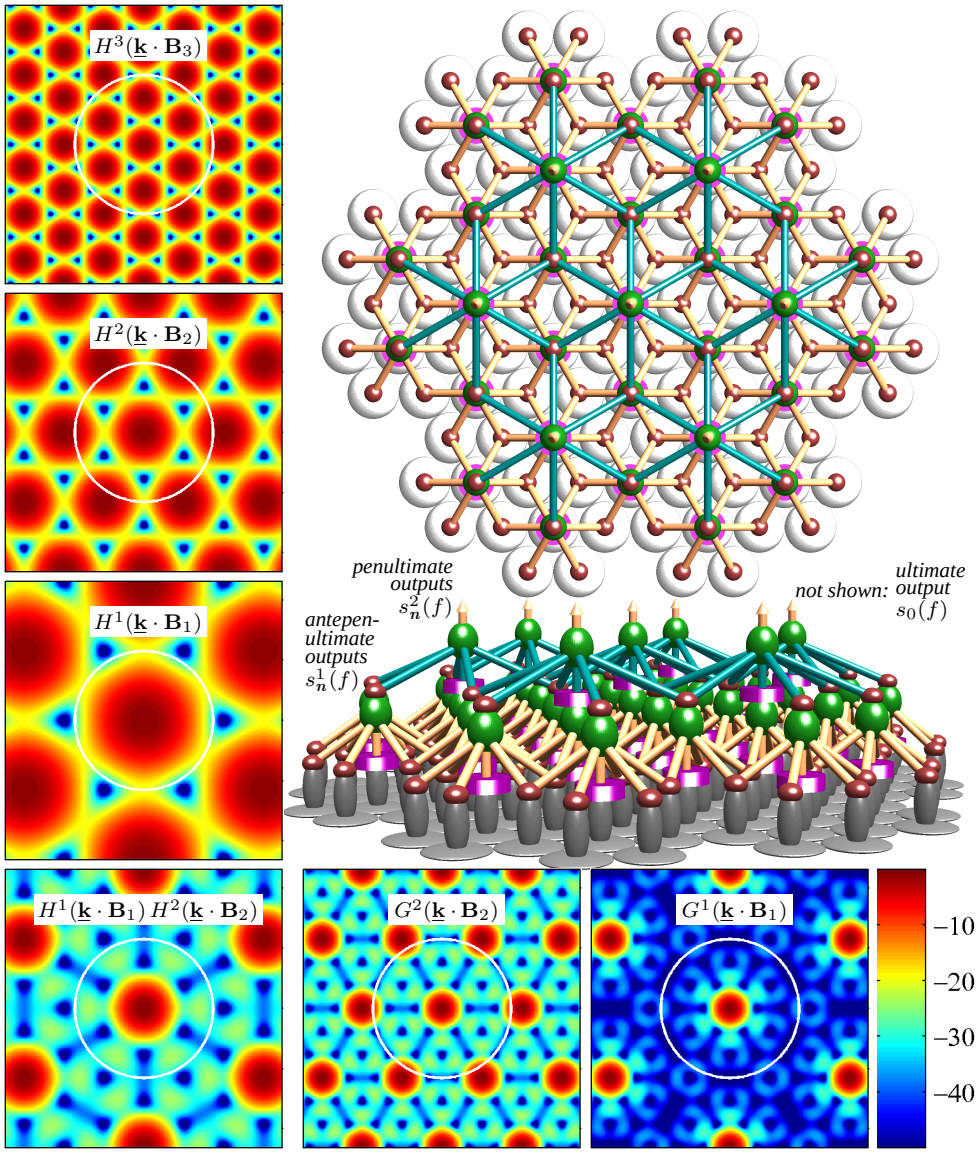

Figure 6: A two-layer subarray system for a 115-element triangular array using $|\mathbf{R}_1| = |\mathbf{R}_2| = 3$ and $|h^1| = |h^2| = |h^3| = 7$.

ultimate output s_0, the computation of which is not shown, and working back toward the elements, each layer's input needs determine which signals the previous layer must provide. Thus the $|h^3| = 7$ penultimate outputs s_n^2 shown are those required by (44) for computing ultimate output s_0. Each of those penultimate outputs s_n^2 in turn requires, through (43), that specific ante-penultimate outputs s_n^1 be available. Some—here away from the edges of the array exactly $2/3$ of them—of those s_n^1 contribute to more than one penultimate output, so the number of ante-penultimate outputs required here is not $|h^2| \times |h^3| = 49$ but only $|g^2| = 31$. Continuing, we finally determine which element outputs s_n^0 must be realized (along with guard elements) and therefore the size of the array. Because many element

outputs are ultimately shared among multiple subarray computations, the array size is not $|h^1|\times|h^2|\times|h^3| = 343$ but only $|g^1| = 115$.

The array-plane locations of the signals. As before, let

$$\text{location of } s^i_{\boldsymbol{n}} = \text{location of } s^{i-1}_{\mathbf{R}_i\boldsymbol{n}}$$

because it is the latter to which coefficient h^i_0 is applied—steering doesn't affect location—when computing $s^i_{\boldsymbol{n}}$ in the convolution in (43) or (44). But the location of $s^0_{\boldsymbol{n}}$ is element location $\mathbf{B}\boldsymbol{n} = \mathbf{B}_1\boldsymbol{n}$, so the recursion and (45) give the location of $s^i_{\boldsymbol{n}}$ more generally as $\mathbf{B}_{i+1}\boldsymbol{n}$. A spatial convolution similar to (42) can then be used to assign location $\mathbf{B}_i\boldsymbol{k}$ to coefficient $h^i_{\boldsymbol{k}}$.

The array factors. To derive the equivalent array factor, substitute (43) with $i = N$ into (44) and apply, right to left, referral identity (22), Noble identity (26), and referral identity (22) again to obtain s as each line in turn:

$$\begin{aligned}
&s^{N-1} \circlearrowleft \underline{\mathbf{k}}_N\cdot\mathbf{B}_N \star h^N \downarrow \mathbf{R}_N \star (\underline{\mathbf{k}}_{N+1}\cdot\mathbf{B}_{N+1} \circlearrowright h^{N+1}) \qquad \circlearrowleft \underline{\mathbf{k}}_{N+1}\cdot\mathbf{B}_{N+1}\\
&s^{N-1} \circlearrowleft \underline{\mathbf{k}}_N\cdot\mathbf{B}_N \star (h^N \star \mathbf{R}_N \uparrow \underline{\mathbf{k}}_{N+1}\cdot\mathbf{B}_{N+1} \circlearrowright h^{N+1}) \qquad \downarrow\mathbf{R}_N \circlearrowleft \underline{\mathbf{k}}_{N+1}\cdot\mathbf{B}_{N+1}\\
&s^{N-1} \star (\underline{\mathbf{k}}_N\cdot\mathbf{B}_N \circlearrowright h^N \star \mathbf{R}_N \uparrow \underline{\mathbf{k}}_{N+1}\cdot\mathbf{B}_{N+1} \circlearrowright h^{N+1}) \circlearrowleft \underline{\mathbf{k}}_N\cdot\mathbf{B}_N \downarrow \mathbf{R}_N \circlearrowleft \underline{\mathbf{k}}_{N+1}\cdot\mathbf{B}_{N+1}.
\end{aligned}$$

Write this as

$$\begin{aligned}
s &= s^{N-1} \star \qquad g^N \qquad \circlearrowleft \underline{\mathbf{k}}_N\cdot\mathbf{B}_N \downarrow \mathbf{R}_N \circlearrowleft \underline{\mathbf{k}}_{N+1}\cdot\mathbf{B}_{N+1}\\
g^N &\triangleq \quad \underline{\mathbf{k}}_N\cdot\mathbf{B}_N \circlearrowright h^N \star \mathbf{R}_N \uparrow \underline{\mathbf{k}}_{N+1}\cdot\mathbf{B}_{N+1} \circlearrowright h^{N+1}.
\end{aligned}$$

Substitute (43) with $i = N-1$ and apply Noble identity (26) and then referral identity (22) and finally (24) to write

$$\begin{aligned}
s &= s^{N-2} \circlearrowleft \underline{\mathbf{k}}_{N-1}\cdot\mathbf{B}_{N-1} \star h^{N-1} \downarrow \mathbf{R}_{N-1} \star g^N \circlearrowleft \underline{\mathbf{k}}_N\cdot\mathbf{B}_N \downarrow \mathbf{R}_N \circlearrowleft \underline{\mathbf{k}}_{N+1}\cdot\mathbf{B}_{N+1}\\
s &= s^{N-2} \star g^{N-1} \circlearrowleft \underline{\mathbf{k}}_{N-1}\cdot\mathbf{B}_{N-1} \downarrow \mathbf{R}_{N-1} \circlearrowleft \underline{\mathbf{k}}_N\cdot\mathbf{B}_N \downarrow \mathbf{R}_N \circlearrowleft \underline{\mathbf{k}}_{N+1}\cdot\mathbf{B}_{N+1}\\
g^{N-1} &\triangleq \underline{\mathbf{k}}_{N-1}\cdot\mathbf{B}_{N-1} \circlearrowright h^{N-1} \star \mathbf{R}_{N-1} \uparrow g^N.
\end{aligned}$$

Continue until the entire input-output relationship is seen to be

$$s \triangleq s^0 \star g^1 \quad \circlearrowleft \underline{\mathbf{k}}_1\cdot\mathbf{B}_1 \downarrow \mathbf{R}_1 \;\cdots\; \circlearrowleft \underline{\mathbf{k}}_N\cdot\mathbf{B}_N \downarrow \mathbf{R}_N \quad \circlearrowleft \underline{\mathbf{k}}_{N+1}\cdot\mathbf{B}_{N+1} \tag{46}$$

$$g^{N+1} \triangleq \underline{\mathbf{k}}_{N+1}\cdot\mathbf{B}_{N+1} \circlearrowright h^{N+1} \tag{47}$$

$$g^i \triangleq \underline{\mathbf{k}}_i\cdot\mathbf{B}_i \circlearrowright h^i \star \mathbf{R}_i \uparrow g^{i+1} \tag{48}$$

for $i \in \{1,\ldots,N\}$. The operations in (46) following the convolution with g^1 do not change s_0, so in determining s_0 those operations can be omitted and (46) replaced with (29), but

now with g^1 determined by (47) and (48). The Fourier transforms of (47) and (48) are, from (21) and (27),

$$G^{N+1}(\boldsymbol{f}) = H^{N+1}(\boldsymbol{f} + \underline{\mathbf{k}}_{N+1} \cdot \mathbf{B}_{N+1}), \tag{49}$$

$$G^{i}(\boldsymbol{f}) = H^{i}(\boldsymbol{f} + \underline{\mathbf{k}}_{i} \cdot \mathbf{B}_{i})\; G^{i+1}((\boldsymbol{f} + \underline{\mathbf{k}}_{i} \cdot \mathbf{B}_{i})\,\mathbf{R}_{i}). \tag{50}$$

Unwrapping recursion (50) to obtain overall array factor $G^1(\underline{\mathbf{k}} \cdot \mathbf{B}_1)$ and then substituting that and $\mathrm{A}(\underline{\mathbf{k}}, f)$ of (13) into (31) gives array output

$$\begin{aligned} s_0(f) = \int \Bigg[\int \overrightarrow{\mathbf{E}}\Big(\big[\underline{\mathbf{k}} + k\widehat{b},\, f/c\big] \Big)\, dk \Bigg] \\ \cdot \Big(\quad \overrightarrow{\underline{\mathbf{\Phi}}}(\underline{\mathbf{k}}, f) \\ \times \quad H^1((\underline{\mathbf{k}} + \underline{\mathbf{k}}_1 \qquad) \cdot \mathbf{B}_1 \quad) \\ \times \quad H^2((\underline{\mathbf{k}} + \underline{\mathbf{k}}_1 + \underline{\mathbf{k}}_2 \quad) \cdot \mathbf{B}_2 \quad) \\ \vdots \\ \times\, H^{N+1}((\underline{\mathbf{k}} + \underline{\mathbf{k}}_1 + \cdots + \underline{\mathbf{k}}_{N+1}) \cdot \mathbf{B}_{N+1}) \Big)\, d\underline{\mathbf{k}}. \end{aligned} \tag{51}$$

The inner integral is the incident array-plane electric field, and directional selectivity is applied to it through the dot product with the $N+2$ functions of $\underline{\mathbf{k}}$ in the large parentheses. Many products of these come up in discussion and so need standard names. The argument lists are suppressed for brevity:

(periodic) element pattern	$\overrightarrow{\underline{\mathbf{\Phi}}}$
subarray i pattern	$\overrightarrow{\underline{\mathbf{\Phi}}}\, H^1 \cdots H^i$
overall array pattern	$\overrightarrow{\underline{\mathbf{\Phi}}}\, H^1 \cdots \quad H^{N+1}$
analog pattern	$\overrightarrow{\underline{\mathbf{\Phi}}}\, H^1 \cdots \quad H^{j-1}$
digital array factor	$H^j \cdots H^{N+1}$
subarray i array factor	H^i
ante-penultimate array factor	H^{N-1}
penultimate array factor	H^N
ultimate array factor	H^{N+1}
overall array factor	$H^1 \quad \cdots H^{N+1}$

Array-factor periodicity, reciprocal bases, and dual lattices. Every frequency response $H^i(\boldsymbol{f})$ has periodic form (17): $H^i(\boldsymbol{f} + \boldsymbol{k}^T) = H^i(\boldsymbol{f})$ for any $\boldsymbol{k} \in \mathbb{Z}^2$. From (49) then, $G^{N+1}(\boldsymbol{f} + \boldsymbol{k}^T) = G^{N+1}(\boldsymbol{f})$, and by (50) and induction on i,

$$\begin{aligned} G^{i}(\boldsymbol{f} + \boldsymbol{k}^T) &= H^{i}(\boldsymbol{f} + \boldsymbol{k}^T + \underline{\mathbf{k}}_{i} \cdot \mathbf{B}_{i})\; G^{i+1}((\boldsymbol{f} + \underline{\mathbf{k}}_{i} \cdot \mathbf{B}_{i})\,\mathbf{R}_{i} + \boldsymbol{k}^T \mathbf{R}_{i}) \\ &= H^{i}(\boldsymbol{f} + \underline{\mathbf{k}}_{i} \cdot \mathbf{B}_{i})\; G^{i+1}((\boldsymbol{f} + \underline{\mathbf{k}}_{i} \cdot \mathbf{B}_{i})\,\mathbf{R}_{i}) \\ &= G^{i}(\boldsymbol{f}), \end{aligned}$$

so $G^i(\boldsymbol{f})$ is similarly periodic.

subarray array factor	*locations of signals applied to*		*passband centers in* $\underline{\mathbf{k}}$
$H^i(\underline{\mathbf{k}} \cdot \mathbf{B}_i)$	$\mathbf{B}_i\mathbb{Z}^2$	$\leftrightarrow$	$\mathbf{B}_i^{-1}\mathbb{Z}^2$
	$\cup$	density ratio $\lvert\mathbf{R}_i\rvert$	$\cap$
$H^{i+1}(\underline{\mathbf{k}} \cdot \mathbf{B}_{i+1})$	$\mathbf{B}_{i+1}\mathbb{Z}^2$	$\leftrightarrow$	$\mathbf{B}_{i+1}^{-1}\mathbb{Z}^2$

Figure 7: Passband locations of the subarray array factors and their relationships.

Array-factor periodicity in $\underline{\mathbf{k}}$ is trickier. From above, $H\big((\underline{\mathbf{k}} + \Delta\underline{\mathbf{k}}) \cdot \mathbf{B}\big) = H(\underline{\mathbf{k}} \cdot \mathbf{B})$ if $\mathbf{B}^T \cdot \Delta\underline{\mathbf{k}} \in \mathbb{Z}^2$. Suppose the latter holds, and let $\boldsymbol{z} = \mathbf{B}^T \cdot \Delta\underline{\mathbf{k}}$. Then $\mathbf{B}^T \cdot \Delta\underline{\mathbf{k}} = \boldsymbol{z} = \mathbf{G}\mathbf{G}^{-1}\boldsymbol{z} = \mathbf{B}^T \cdot \mathbf{B}\mathbf{G}^{-1}\boldsymbol{z}$, where basis $\mathbf{B}$ has Gram matrix $\mathbf{G} = \mathbf{B}^T \cdot \mathbf{B}$ as before. The dot products of $\Delta\underline{\mathbf{k}}$ and $\mathbf{B}\mathbf{G}^{-1}\boldsymbol{z}$ with the two components of $\mathbf{B}^T = \left[\begin{smallmatrix}\mathbf{b}_1\\ \mathbf{b}_2\end{smallmatrix}\right]$ are equal, so $\Delta\underline{\mathbf{k}} = \mathbf{B}\mathbf{G}^{-1}\boldsymbol{z}$ by the linear independence of $\mathbf{b}_1$ and $\mathbf{b}_2$. Here $\boldsymbol{z} \in \mathbb{Z}^2$ is arbitrary, so we have our periodicity condition: $H\big((\underline{\mathbf{k}} + \Delta\underline{\mathbf{k}}) \cdot \mathbf{B}\big) = H(\underline{\mathbf{k}} \cdot \mathbf{B})$ if $\Delta\underline{\mathbf{k}} \in \mathbf{B}^{-1}\mathbb{Z}^2$, where $\mathbf{B}^{-1} \triangleq \mathbf{B}\mathbf{G}^{-1}$ is the *reciprocal*[1] of basis $\mathbf{B}$. Lattice $\mathbf{B}^{-1}\mathbb{Z}^2$ is the *dual* of lattice $\mathbf{B}\mathbb{Z}^2$. Reciprocal basis $\mathbf{B}^{-1}$ has Gram matrix $\mathbf{G}^{-1}\mathbf{B}^T \cdot \mathbf{B}\mathbf{G}^{-1} = \mathbf{G}^{-1}$, the inverse of the original Gram matrix, so the reciprocal of reciprocal basis $\mathbf{B}^{-1}$ is $\mathbf{B}\mathbf{G}^{-1}\mathbf{G} = \mathbf{B}$ and the dual of dual lattice $\mathbf{B}^{-1}\mathbb{Z}^2$ is $\mathbf{B}\mathbb{Z}^2$ so that each lattice is the dual of the other. Following signal-processing custom, by a *period* of array factor $H(\underline{\mathbf{k}} \cdot \mathbf{B})$ we'll mean a period of dual lattice $\mathbf{B}^{-1}\mathbb{Z}^2$, and by *the* period we'll mean an arbitrary period.

Sublattice basis $\mathbf{B}\mathbf{R}$ has Gram matrix $\mathbf{R}^T\mathbf{B}^T \cdot \mathbf{B}\mathbf{R} = \mathbf{R}^T\mathbf{G}\mathbf{R}$, and its reciprocal is therefore $\mathbf{B}\mathbf{R}\big(\mathbf{R}^T\mathbf{G}\mathbf{R}\big)^{-1} = \mathbf{B}\mathbf{G}^{-1}\mathbf{R}^{-T} = \mathbf{B}^{-1}\mathbf{R}^{-T}$. Its Gram matrix $\mathbf{R}^{-1}\mathbf{G}^{-1}\mathbf{B}^T \cdot \mathbf{B}\mathbf{G}^{-1}\mathbf{R}^{-T} = \mathbf{R}^{-1}\mathbf{G}^{-1}\mathbf{R}^{-T}$ gives the period of dual lattice $\mathbf{B}^{-1}\mathbf{R}^{-T}\mathbb{Z}^2$ an area of $\lvert\mathbf{R}^{-1}\mathbf{G}^{-1}\mathbf{R}^{-T}\rvert^{1/2} = \lvert\mathbf{G}\rvert^{-1/2}\lvert\mathbf{R}\rvert^{-1}$. In this diagram $\big(\mathbf{B}^{-1}\mathbf{R}^{-T}\big)\mathbf{R}^T\mathbb{Z}^2 = \mathbf{B}^{-1}\mathbb{Z}^2$ yields the rightmost sublattice relationship:

$$\begin{array}{ccc} \text{lattice } \mathbf{B}\mathbb{Z}^2 & \overset{\text{duals}}{\leftrightarrow} & \text{dual } \mathbf{B}^{-1}\mathbb{Z}^2 \text{ of lattice} \\ \text{(density ratio } \lvert\mathbf{R}\rvert) \quad \cup & & \cap \quad \text{(density ratio } \lvert\mathbf{R}\rvert) \\ \text{sublattice } \mathbf{B}\mathbf{R}\mathbb{Z}^2 & \overset{\text{duals}}{\leftrightarrow} & \text{dual } \mathbf{B}^{-1}\mathbf{R}^{-T}\mathbb{Z}^2 \text{ of sublattice} \end{array}$$

The rôle of subarray array factors. Specializing the above sublattice relationship to the array factors of (51) reveals the Fig. 7 relationships, visible in the Fig. 6 example. Each unsteered subarray array factor $H^i(\underline{\mathbf{k}} \cdot \mathbf{B}_i)$ must have a passband centered at $\underline{\mathbf{k}} = 0$ so that unsteered overall array factor $G^1(\underline{\mathbf{k}} \cdot \mathbf{B})$ has such a passband, the array "main beam." By

[1] Be careful of this notation. In this paper a basis $\mathbf{B}$ is not a matrix, and $\mathbf{B}^{-1}$ does not denote a matrix inverse. One often represents a basis $\mathbf{B}$ computationally as $\mathbf{B} = \left[\begin{smallmatrix}\widehat{\imath} & \widehat{\jmath} & \widehat{k}\end{smallmatrix}\right]\mathbf{B}_c$ using a matrix $\mathbf{B}_c$ of coordinates applied to Cartesian unit vectors. (This $\widehat{k}$ is of course distinct from the $\mathbf{k}$-related $\widehat{k}$ used earlier.) Then $\mathbf{G} = \mathbf{B}_c^T\mathbf{B}_c$, and $\mathbf{B}^{-1} = \left[\begin{smallmatrix}\widehat{\imath} & \widehat{\jmath} & \widehat{k}\end{smallmatrix}\right]\mathbf{B}_c\mathbf{G}^{-1}$, so the Cartesian-coordinate representation of the reciprocal basis is $\mathbf{B}_c\mathbf{G}^{-1}$, where now $\mathbf{G} = \mathbf{B}_c^T\mathbf{B}_c$. If we reduce this to two spatial dimensions by eliminating $\widehat{k}$ and making Cartesian-coordinate matrix $\mathbf{B}_c$ only 2×2, then $\mathbf{B}_c\mathbf{G}^{-1} = \mathbf{B}_c(\mathbf{B}_c^T\mathbf{B}_c)^{-1} = \mathbf{B}_c^{-T}$. That transpose is important and is why the basis and its coordinate representation must not be conflated.

periodicity $H^i(\underline{\mathbf{k}} \cdot \mathbf{B}_i)$ then has additional passbands centered on the points of dual lattice $\mathbf{B}_i^{-1}\mathbb{Z}^2$, so the density in the $\underline{\mathbf{k}}$ plane of its passbands is the reciprocal $|\mathbf{G}|^{1/2}\,|\mathbf{R}_1|\cdots|\mathbf{R}_{i-1}|$ of the area of the period of that dual, and each period of subarray array factor $H^i(\underline{\mathbf{k}} \cdot \mathbf{B}_i)$ contains exactly $|\mathbf{R}_i|$ passbands of $H^{i+1}(\underline{\mathbf{k}} \cdot \mathbf{B}_{i+1})$. For $i \leq N$ then, $H^i(\underline{\mathbf{k}} \cdot \mathbf{B}_i)$ functions as a masking filter that passes one of these while masking out the other $|\mathbf{R}_i| - 1$ of them so that those in the visible region do not become grating lobes.

Resolution and the rôle of the ultimate array factor. FIR-filter frequency response $H(\boldsymbol{f})$ has $|h|$ design degrees of freedom available per $\boldsymbol{f}^T$ period, the unit square, so on average each coefficient controls an area of $|h|^{-1}$. That unit square is mapped to each period of array factor $H(\underline{\mathbf{k}} \cdot \mathbf{BR})$, so each coefficient effectively controls $H(\underline{\mathbf{k}} \cdot \mathbf{BR})$ over an area in $\underline{\mathbf{k}}$ of $(L\,|\mathbf{G}|^{1/2}\,|\mathbf{R}|)^{-1}$, the *resolution* of $H(\underline{\mathbf{k}} \cdot \mathbf{BR})$. The resolution of array factor $H^i(\underline{\mathbf{k}} \cdot \mathbf{B}_i)$ in particular then decreases with i, so that ultimate array factor $H^{N+1}(\underline{\mathbf{k}} \cdot \mathbf{B}_{N+1})$ has the smallest resolution limit. We always use the latter then as the shaping filter that determines the precise shape of the main beam, the $\underline{\mathbf{k}} = 0$ passband in unsteered overall array factor $G^1(\underline{\mathbf{k}} \cdot \mathbf{B})$.

Specification of individual array factors. The $\underline{\mathbf{k}} = 0$ passband of unsteered ultimate array factor $H^{N+1}(\underline{\mathbf{k}} \cdot \mathbf{B}_{N+1})$, the shaping filter, is a set $\mathcal{K}_p$ of $\underline{\mathbf{k}}$ vectors over which its performance is adequate to ensure adequate communication, target detection, etc. as per the array's purpose. In a volume-search radar for example, adequacy means a sort of net passband coverage of the search area, the portion of the visible region of interest, by the steered ultimate array factor in (51). More precisely, the *scanned passband* set

$$\bigcup_{(\underline{\mathbf{k}}_1 \ldots \underline{\mathbf{k}}_{N+1}) \in \substack{\text{combinations} \\ \text{stepped through}}} \mathcal{K}_p + \underline{\mathbf{k}}_1 + \cdots + \underline{\mathbf{k}}_{N+1}$$

must contain the search area. Typically $\underline{\mathbf{k}}_1 + \cdots + \underline{\mathbf{k}}_{N+1}$ steps through some lattice of which $\mathcal{K}_p$ is a period. That period might be, give or take the inclusion or exclusion of edges, the lattice's *Voronoi region*, the set of $\underline{\mathbf{k}}$ for which the origin is the nearest lattice point. The complete passband set of $H^{N+1}(\underline{\mathbf{k}} \cdot \mathbf{B}_{N+1})$, including periodically repeated passbands, is $\mathcal{K}_p + \mathbf{B}_{N+1}^{-1}\mathbb{Z}^2$.

Define some set $\mathcal{K}_s$ to include both $\mathcal{K}_p$ and a surrounding transition region. The complete stopband set of ultimate array factor $H^{N+1}(\underline{\mathbf{k}} \cdot \mathbf{B}_{N+1})$ can then be written as the complement $\left(\mathcal{K}_s + \mathbf{B}_{N+1}^{-1}\mathbb{Z}^2\right)^c$ of its periodic repeats. For $\underline{\mathbf{k}}$ in this set, $|H^{N+1}(\underline{\mathbf{k}} \cdot \mathbf{B}_{N+1})|$ must be reduced to sidelobe levels. In the designs in this paper $\mathcal{K}_s$ is always circular and centered on the origin, but there are special cases in which other shapes are advantageous, particularly square shapes when $\mathbf{B} = \mathbf{B}_\square$ and $\mathbf{R} = R\mathbf{I}$.

The complete passband set for masking filter $H^i(\underline{\mathbf{k}} \cdot \mathbf{B}_i)$ is $\mathcal{K}_p + \mathbf{B}_i^{-1}\mathbb{Z}^2$. Its stopband centers comprise set difference $\mathbf{B}_{i+1}^{-1}\mathbb{Z}^2 \setminus \mathbf{B}_i^{-1}\mathbb{Z}^2$, so its complete stopband set is $\mathcal{K}_s + \mathbf{B}_{i+1}^{-1}\mathbb{Z}^2 \setminus \mathbf{B}_i^{-1}\mathbb{Z}^2$.

3.2. Filter Design

Here we address the design of coefficient sets $h^1, \ldots, h^{N+1}$ for spatial filtering in multi-layer subarray systems. We focus on structural approaches to simplifying both the filters and the formulations through which they are optimized and omit such formulation details as are not specific to subarray systems, both because they are covered adequately elsewhere [6, 7]. and because they would be irrelevant here: there are many reasonable ways to optimize filters with such small numbers of distinct coefficient values as these, including, for the smaller filters, simple trial and error. For the record, our optimizations used a matlab-based tool for linear- and second-order cone programming [8].

Lattice symmetries. An *orthogonal* transformation of the array plane is a linear map $\mathcal{L}$ from the array plane into itself that preserves dot products, so that $(\mathcal{L}\mathbf{x})\cdot(\mathcal{L}\mathbf{y}) = \mathbf{x}\cdot\mathbf{y}$ for any array-plane vectors $\mathbf{x}$ and $\mathbf{y}$. An orthogonal transformation $\mathcal{L}$ is a *symmetry* of lattice $\mathbf{B}\mathbb{Z}^2$ if the latter is closed under $\mathcal{L}$.

A lattice symmetry can be represented using a specific lattice basis. If $\mathcal{L}$ is a linear array-plane operator, then it is completely characterized by its action on any two linearly independent vectors in the array plane. Let's use the two basis vectors in $\mathbf{B}$ and express the vectors to which they map in that same basis by writing $\mathcal{L}\mathbf{B} = \mathbf{B}\mathbf{S}$, a convenient shorthand for $\mathcal{L}\mathbf{b}_1 = s_{11}\mathbf{b}_1 + s_{21}\mathbf{b}_2$ and $\mathcal{L}\mathbf{b}_2 = s_{12}\mathbf{b}_1 + s_{22}\mathbf{b}_2$, where $\mathbf{S}$ is a 2×2 matrix that now represents $\mathcal{L}$. If lattice $\mathbf{B}\mathbb{Z}^2$ is to be closed under $\mathcal{L}$, then for every $\boldsymbol{z} \in \mathbb{Z}^2$ we need $\mathcal{L}(\mathbf{B}\boldsymbol{z}) = (\mathcal{L}\mathbf{B})\boldsymbol{z} = \mathbf{B}\mathbf{S}\boldsymbol{z} \in \mathbf{B}\mathbb{Z}^2$, which is so if and only if $\mathbf{S}$ is an integer matrix. Further, map $\mathcal{L}$ is orthogonal if for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^2$ the quantity $(\mathbf{B}\boldsymbol{x}) \cdot (\mathbf{B}\boldsymbol{y}) = \boldsymbol{x}^T\mathbf{B}^T \cdot \mathbf{B}\boldsymbol{y} = \boldsymbol{x}^T\mathbf{G}\boldsymbol{y}$, where $\mathbf{G} = \mathbf{B}^T \cdot \mathbf{B}$ is the Gram matrix of basis $\mathbf{B}$, is the same as the quantity $(\mathcal{L}(\mathbf{B}\boldsymbol{x})) \cdot (\mathcal{L}(\mathbf{B}\boldsymbol{y}) = ((\mathcal{L}\mathbf{B})\boldsymbol{x}) \cdot ((\mathcal{L}\mathbf{B})\boldsymbol{y}) = (\mathbf{B}\mathbf{S}\boldsymbol{x}) \cdot (\mathbf{B}\mathbf{S}\boldsymbol{y}) = \boldsymbol{x}^T\mathbf{S}^T\mathbf{B}^T \cdot \mathbf{B}\mathbf{S}\boldsymbol{y} = \boldsymbol{x}^T\mathbf{S}^T\mathbf{G}\mathbf{S}\boldsymbol{y}$. These two quantities are identical for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^2$ if and only if $\mathbf{S}^T\mathbf{G}\mathbf{S} = \mathbf{G}$, that is, if the operator preserves the Gram matrix of the lattice. Given any lattice symmetry $\mathcal{L}$, the matrix $\mathbf{S}$ that represents it in this way is basis dependent, so matrix $\mathbf{S}$ is a *symmetry of lattice basis* $\mathbf{B}$.

Given a lattice symmetry $\mathcal{L}$, there is always a 2×2 integer matrix $\mathbf{S}$ such that $\mathcal{L}\mathbf{B} = \mathbf{B}\mathbf{S}$, a functional relationship we can write as $\phi(\mathcal{L}) = \mathbf{S}$. Suppose $\phi(\mathcal{L}_1) = \mathbf{S}_1$ and $\phi(\mathcal{L}_2) = \mathbf{S}_2$. If we apply $\mathcal{L}_1$ and then $\mathcal{L}_2$, then we've mapped $\mathbf{B}\boldsymbol{z} \mapsto \mathbf{B}\mathbf{S}_1\boldsymbol{z} \mapsto \mathbf{B}\mathbf{S}_2\mathbf{S}_1\boldsymbol{z}$, establishing that matrix $\mathbf{S}_2\mathbf{S}_1$ represents symmetry $\mathcal{L}_2\mathcal{L}_1$, that $\phi(\mathcal{L}_2\mathcal{L}_1) = \mathbf{S}_2\mathbf{S}_1$. This sequential-application operator is associative, and its identity element is the identity map specified by the identity matrix. A lattice symmetry will always have an inverse with respect to this identity, so the symmetries of a lattice will always form a group under this operator, the *symmetry group* $\mathcal{S}$ of the lattice. Property $\phi(\mathcal{L}_2\mathcal{L}_1) = \mathbf{S}_2\mathbf{S}_1$ of map ϕ (making ϕ a homomorphism) ensures that that the range of ϕ, the set $\mathcal{M}$ of matrices representing the symmetries of $\mathcal{S}$, forms a group under matrix multiplication, the *symmetry group of the lattice basis*. There is only one $\mathcal{L}$ for which $\phi(\mathcal{L}) = \mathbf{I}$, so there is a one-to-one correspondence between the subgroups of $\mathcal{S}$ and the subgroups of $\mathcal{M}$ (i.e. $\mathcal{S}$ and $\mathcal{M}$ are isomorphic).

Suppose the Gram matrix of $\mathbf{B}$ has form $\mathbf{G} = \frac{d^2}{2}\left[\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right]$. This is so for $\mathbf{B}_\triangle$, for $\mathbf{B}_\triangle\mathbf{R}$ where $\mathbf{R}$ is one or more of the $\mathbf{B}_\triangle$ resampling matrices in Fig. 2, and for all their reciprocals.

The symmetries of $\mathbf{B}$ are these matrices:

$\mathbf{S}_{00} \triangleq \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ identity	$\mathbf{S}_{01} \triangleq \begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix}$ rotation by $60°$	$\mathbf{S}_{02} \triangleq \begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix}$ rotation by $120°$	$\mathbf{S}_{03} \triangleq \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$ rotation by $180°$	$\mathbf{S}_{04} \triangleq \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix}$ rotation by $240°$	$\mathbf{S}_{05} \triangleq \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix}$ rotation by $300°$
$\mathbf{S}_{10} \triangleq \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ reflection in $30°$ line	$\mathbf{S}_{11} \triangleq \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}$ reflection in $0°$ line	$\mathbf{S}_{12} \triangleq \begin{bmatrix} 1 & 0 \\ -1 & -1 \end{bmatrix}$ reflection in $150°$ line	$\mathbf{S}_{13} \triangleq \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$ reflection in $120°$ line	$\mathbf{S}_{14} \triangleq \begin{bmatrix} -1 & -1 \\ 0 & 1 \end{bmatrix}$ reflection in $90°$ line	$\mathbf{S}_{15} \triangleq \begin{bmatrix} -1 & 0 \\ 1 & 1 \end{bmatrix}$ reflection in $60°$ line

The angles begin in the direction of basis vector $\mathbf{b}_1$ and increase toward basis vector $\mathbf{b}_2$. (This $\mathbf{G}$ specifies these basis vectors to be $60°$ apart.) The six rotations alone form the *rotation group* $\mathcal{R}$ of the lattice. Rotation group $\mathcal{R}$ is always a subgroup of $\mathcal{S}$. The other members of $\mathcal{S}$ are reflections.

Symmetry in filters and array factors. Let $\mathcal{G}$ be any finite group of 2×2 integer matrices under matrix multiplication. Let $\mathbf{S} \in \mathcal{G}$ be arbitrary. Certainly $|\mathbf{S}| = 1$, because otherwise either $|\mathbf{S}| = 0$, so that $\mathbf{S}$ has no inverse and $\mathcal{G}$ is not a group, or for the various $n > 0$ both $|\mathbf{S}|^n$ and (therefore) $\mathbf{S}^n$ are distinct and group $\mathcal{G}$ is infinite. But $|\mathbf{S}| = 1$ makes map $\boldsymbol{z} \mapsto \mathbf{S}\boldsymbol{z}$ an automorphism of $\mathbb{Z}^2$, so a sum indexed over $\boldsymbol{n} \in \mathbb{Z}^2$ can be turned into one indexed over $\boldsymbol{m} \in \mathbb{Z}^2$ with change of index $\boldsymbol{m} = \mathbf{S}\boldsymbol{n}$. The same terms are summed. In Fourier pair $h_{\boldsymbol{n}} \leftrightarrow H(\boldsymbol{f}) = \sum_{\boldsymbol{n}} h_{\boldsymbol{n}}\, \mathrm{e}^{-\mathrm{j}2\pi \boldsymbol{f}\boldsymbol{n}}$ then, replace each coefficient $h_{\boldsymbol{n}}$ with $h_{\mathbf{S}\boldsymbol{n}}$ to obtain $h_{\mathbf{S}\boldsymbol{n}} \leftrightarrow \sum_{\boldsymbol{n}} h_{\mathbf{S}\boldsymbol{n}} \mathrm{e}^{-\mathrm{j}2\pi \boldsymbol{f}\boldsymbol{n}} = \sum_{\boldsymbol{m}} h_{\boldsymbol{m}} \mathrm{e}^{-\mathrm{j}2\pi \boldsymbol{f}\mathbf{S}^{-1}\boldsymbol{m}} = H(\boldsymbol{f}\mathbf{S}^{-1})$. Now suppose coefficient set h is *left invariant under* $\mathcal{G}$ in the sense that for an arbitrary $\mathbf{S} \in \mathcal{G}$, coefficient $h_{\boldsymbol{n}} = h_{\mathbf{S}\boldsymbol{n}}$ for every $\boldsymbol{n} \in \mathbb{Z}^2$. As functions of $\boldsymbol{f}$ then, $H(\boldsymbol{f}) = H(\boldsymbol{f}\mathbf{S}^{-1})$ or, equivalently, $H(\boldsymbol{f}\mathbf{S}) = H(\boldsymbol{f})$. Here $\mathbf{S} \in \mathcal{G}$ is arbitrary, so $H(\boldsymbol{f})$ is *right invariant under* $\mathcal{G}$. Conversely, if $H(\boldsymbol{f})$ is right invariant under $\mathcal{G}$ and $\mathbf{S} \in \mathcal{G}$ then $H(\boldsymbol{f}) = H(\boldsymbol{f}\mathbf{S}^{-1})$ implies $h_{\boldsymbol{n}} = h_{\mathbf{S}\boldsymbol{n}}$, so h is left invariant under $\mathcal{G}$. Left and right invariance under $\mathcal{G}$ on the left and right respectively of Fourier pair $h_{\boldsymbol{n}} \leftrightarrow H(\boldsymbol{f})$ are equivalent.

Before considering array factors, we'll need a quick intermediate result. Let $h \leftrightarrow H$ be a Fourier pair. For any orthogonal transformation $\mathcal{L}$, any invertible 2×2 matrix $\mathbf{S}$, and any $\underline{\mathbf{k}}$ in the array plane, $H(\underline{\mathbf{k}} \cdot \mathbf{B}\mathbf{S}^{-1}) = H\big(\mathcal{L}\underline{\mathbf{k}} \cdot \mathcal{L}(\mathbf{B}\mathbf{S}^{-1})\big) = H\big(\mathcal{L}\underline{\mathbf{k}} \cdot \mathcal{L}(\mathbf{B})\mathbf{S}^{-1}\big) = H(\mathcal{L}\underline{\mathbf{k}} \cdot \mathbf{B}\mathbf{S}\mathbf{S}^{-1}) = H(\mathcal{L}\underline{\mathbf{k}} \cdot \mathbf{B})$. We can use this to enable a pair of converse relationships.

First, let $\mathcal{G}$ be a subgroup (or all) of $\mathcal{M}$, and let $\mathbf{B}\mathcal{G}$ denote ($\phi^{-1}(\mathcal{G})$, which is) the family of linear operators $\mathcal{L}$ defined from an $\mathbf{S} \in \mathcal{G}$ by $\mathcal{L}\mathbf{B} = \mathbf{B}\mathbf{S}$. Let $\mathcal{L} \in \mathbf{B}\mathcal{G}$ be arbitrary,and suppose $\mathcal{L}\mathbf{B} = \mathbf{B}\mathbf{S}$ and $\mathbf{S} \in \mathcal{G}$. Then $\mathbf{S}$ is invertible, and Suppose $h \leftrightarrow H$ is invariant under $\mathcal{G}$. Then $H(\boldsymbol{f}) = H(\boldsymbol{f}\mathbf{S}^{-1})$ for all $\boldsymbol{f}$, so $H(\underline{\mathbf{k}} \cdot \mathbf{B}) = H(\underline{\mathbf{k}} \cdot \mathbf{B}\mathbf{S}^{-1}) = H(\mathcal{L}\underline{\mathbf{k}} \cdot \mathbf{B})$ by the quick intermediate result above. As functions of $\underline{\mathbf{k}}$ then, $H(\underline{\mathbf{k}} \cdot \mathbf{B}) = H(\mathcal{L}\underline{\mathbf{k}} \cdot \mathbf{B})$ for an arbitrary $\mathcal{L} \in \mathbf{B}\mathcal{G}$, so we term array factor $H(\underline{\mathbf{k}} \cdot \mathbf{B})$ *symmetric with respect to* $\mathbf{B}\mathcal{G}$.

Conversely, suppose array factor $H(\underline{\mathbf{k}} \cdot \mathbf{B})$ is symmetric with respect to $\mathbf{B}\mathcal{G}$, and let $\mathbf{S} \in \mathcal{G}$ be arbitrary. Then $\mathbf{S}^{-1} \in \mathcal{G}$, and $\mathcal{L}\mathbf{B} = \mathbf{B}\mathbf{S}$ defines an $\mathcal{L} \in \mathbf{B}\mathcal{G}$. For any real two-element row vector $\boldsymbol{f}$ there is a vector $\underline{\mathbf{k}}$ with $\boldsymbol{f}\mathbf{S} = \underline{\mathbf{k}} \cdot \mathbf{B}$ by the linear independence of $\mathbf{b}_1$ and $\mathbf{b}_2$. But

$H(\underline{\mathbf{k}}\cdot\mathbf{B})$ is symmetric with respect to $\mathbf{B}\mathcal{G}$, so $H(\underline{\mathbf{k}}\cdot\mathbf{B}) = H(\mathcal{L}\underline{\mathbf{k}}\cdot\mathbf{B}) = H(\underline{\mathbf{k}}\cdot\mathbf{B}\mathbf{S}^{-1})$ by the quick intermediate result above, so $H(\boldsymbol{f}\mathbf{S}) = H(\boldsymbol{f})$. Since $\boldsymbol{f}$ and $\mathbf{S}$ were arbitrary, $h \leftrightarrow H$ is invariant under $\mathcal{G}$, which condition is therefore seen to be equivalent to the symmetry of array factor $H(\underline{\mathbf{k}}\cdot\mathbf{B})$ with respect to $\mathbf{B}\mathcal{G}$.

Imposing symmetry. We have established something very useful: If $\mathcal{G}$ and $\mathbf{B}\mathcal{G}$ are corresponding subgroups of $\mathcal{M}$ and $\mathcal{S}$, the left invariance of h under $\mathcal{G}$ and the symmetry of array factor $H(\underline{\mathbf{k}}\cdot\mathbf{B})$ with respect to $\mathbf{B}\mathcal{G}$ are equivalent. Typically, this array-factor symmetry is completely innocuous, and insisting on it is harmless. All of the hardware array factors in Figs. 1, 3, 6, and 10 are symmetric with respect to the twelve-member example symmetry group $\mathcal{S}$ presented above, and this harmless symmetry gives us left invariance of the corresponding coefficient sets h under $\mathcal{G}$. Such invariance can often save a significant amount of hardware by permitting identical coefficients to be factored from several terms in the first sum in (16). Those terms can be summed and the coefficient applied just once instead of several times. Using one common coefficient in place of several individual ones can also speed up coefficient optimization significantly.

To discover how to use this, let us first return to our 12-member example symmetry group $\mathcal{S}$ above. By $\sigma(\mathbf{S}_{01})$ we mean the smallest subgroup of $\mathcal{M}$ containing $\mathbf{S}_{01}$. A group containing matrix $\mathbf{S}_{01}$ must contain its powers also, so $\sigma(\mathbf{S}_{01}) = \phi(\mathcal{R})$. Similarly, $\sigma(\mathbf{S}_{01}, \mathbf{S}_{10})$ is the smallest subgroup of $\mathcal{S}$ that contains both $\mathbf{S}_{01}$ and $\mathbf{S}_{10}$, so $\sigma(\mathbf{S}_{01}, \mathbf{S}_{10}) = \mathcal{M}$. We say $\mathbf{S}_{01}$ and $\mathbf{S}_{10}$ *generate* $\mathcal{M}$ or $\mathbf{S}_{01}$ and $\mathbf{S}_{10}$ are *generators* for $\mathcal{M}$. In this example, the combination of $\mathbf{S}_{01}$ or $\mathbf{S}_{04}$ and the matrix $\mathbf{S}_{1i}$ representing any reflection generates $\mathcal{M}$. An arbitrary $\mathbf{S} \in \mathcal{M}$ can be written as the product of generators of $\mathcal{M}$. Here $\mathbf{S} = \mathbf{S}_{10}^{j}\mathbf{S}_{01}^{k}$ for some $j \in \{0, 1\}$ and some $k \in \{0, \ldots, 5\}$.

Can we construct h to be left invariant under some subgroup $\mathcal{G}$ of $\mathcal{M}$? Suppose we are permitting $h_{\boldsymbol{n}} \neq 0$ for $\boldsymbol{n} \in \mathcal{Z} \subset \mathbb{Z}^2$ and defaulting to $h_{\boldsymbol{n}} = 0$ for other $\boldsymbol{n}$. Construct $h_{\boldsymbol{n}} = \sum_{\mathbf{S}'\in\mathcal{G}} h'_{\mathbf{S}'\boldsymbol{n}}$ using some intermediate coefficients h' and (in practice) a small set of generators for $\mathcal{G}$ to obtain each $\mathbf{S}'$ needed. Left invariance under $\mathcal{G}$ follows from $h_{\mathbf{S}\boldsymbol{n}} = \sum_{\mathbf{S}'\in\mathcal{G}} h'_{\mathbf{S}'\mathbf{S}\boldsymbol{n}} = \sum_{\mathbf{S}''\in\mathcal{G}} h'_{\mathbf{S}''\boldsymbol{n}} = h_{\boldsymbol{n}}$, where change of index $\mathbf{S}'' \triangleq \mathbf{S}'\mathbf{S}$ sums the same terms because inverting the map $\mathbf{S}' \mapsto \mathbf{S}'\mathbf{S}$ on $\mathcal{G}$ so defined yields a unique $\mathbf{S}''\mathbf{S}^{-1} \in \mathcal{G}$ for any $\mathbf{S}'' \in \mathcal{G}$ and so makes that map one-to-one and onto $\mathcal{G}$. Coefficient locations $\mathbf{B}\mathcal{Z}$ are depicted in Fig. 8 using $\mathbf{B} = \mathbf{B}_{\triangle}$ and five $\mathcal{Z}$ choices with points in $\mathbf{B}\mathcal{Z}_0$ shown lighter, where each set $\mathcal{Z}_0 \subset \mathcal{Z}$ of *launch vectors* is such that their *orbit* $\mathcal{G}\mathcal{Z}_0 \triangleq \bigcup_{\mathbf{S}'\in\mathcal{G}} \mathbf{S}'\mathcal{Z}_0$ *under the action of group* $\mathcal{G}$ is all of $\mathcal{Z}$. Each $h'_{\boldsymbol{n}}$ can now be fixed (in circumstances discussed below) or set to a unique optimization variable as appropriate for each $\boldsymbol{n} \in \mathcal{Z}_0 \subset \mathcal{Z}$ and set to zero for other $\boldsymbol{n}$. Jointly optimizing those variables subject to performance requirements on h optimizes the latter indirectly. It is tedious to prove, but this construction introduces no unused degrees of freedom into the optimization if, as in each Fig. 8 case, set $\mathcal{Z}_0$ is minimal in the sense that removing any point of $\mathcal{Z}_0$ causes $\mathcal{G}\mathcal{Z}_0 = \mathcal{Z}$ to fail.

Spatial Nth-band filters. Imposing symmetries on filter coefficients simplified hardware above without harming array factors. In any masking filter h^i for $i \in \{1, \ldots, N\}$ we'll gain

further (mostly) harmless hardware savings here by forcing $1/|\mathbf{R}_i|$ of the $\boldsymbol{n} \neq 0$ coefficients to zero. This eliminates many splitter outputs, combiner inputs, and transmission lines.

Each period of $H^i(\underline{\mathbf{k}} \cdot \mathbf{B}_i)$ contains one passband and $|\mathbf{R_i}| - 1$ stopbands, both aligned with passbands of $H^{i+1}(\underline{\mathbf{k}} \cdot \mathbf{B}_{i+1})$, so per Fig. 7 each of these $H^i(\underline{\mathbf{k}} \cdot \mathbf{B}_i)$ bands of interest is centered on some point $\mathbf{B}_{i+1}^{-1}\boldsymbol{p}$ with $\boldsymbol{p} \in \mathbb{Z}^2$. Arbitrarily choose some period of $H^i(\underline{\mathbf{k}} \cdot \mathbf{B}_i)$ as *the* period, and let the family of *band designators* denoted by $\left[\mathbb{Z}^2/\mathbf{R}_i^T\mathbb{Z}^2\right]$ (notation from elementary group theory) contain the $\boldsymbol{p}$ vectors for those $|\mathbf{R}_i|$ bands of interest that have centers in the period. The full set of passband centers for $H^i(\underline{\mathbf{k}} \cdot \mathbf{B}_i)$ is then given by $\mathbf{B}_{i+1}^{-1}\boldsymbol{p}_0 + \mathbf{B}_i^{-1}\mathbb{Z}^2$ for some $\boldsymbol{p}_0 \in \left[\mathbb{Z}^2/\mathbf{R}_i^T\mathbb{Z}^2\right]$ through periodic translation of the period's one passband center $\mathbf{B}_{i+1}^{-1}\boldsymbol{p}_0$. Similarly, for any other $\boldsymbol{p}_s \in \left[\mathbb{Z}^2/\mathbf{R}_i^T\mathbb{Z}^2\right]$, stopband center $\mathbf{B}_{i+1}^{-1}\boldsymbol{p}_s$ is in the period and $\mathbf{B}_{i+1}^{-1}\boldsymbol{p}_s + \mathbf{B}_i^{-1}\mathbb{Z}^2$ is the set of stopband centers related to it through periodic translation. Every band of interest in all of $H^i(\underline{\mathbf{k}} \cdot \mathbf{B}_i)$ is related through periodic translation in this way to just one band of interest in the period, so

$$\mathbf{B}_{i+1}^{-1}\mathbb{Z}^2 = \overset{\text{disjoint}}{\bigcup_{\boldsymbol{p} \in [\mathbb{Z}^2/\mathbf{R}_i^T\mathbb{Z}^2]}} \mathbf{B}_{i+1}^{-1}\boldsymbol{p} + \mathbf{B}_i^{-1}\mathbb{Z}^2. \tag{52}$$

Aside: Translating each side by $\mathbf{B}_{i+1}^{-1}\boldsymbol{p}'$ changes nothing fundamental, because on the left $\mathbf{B}_{i+1}^{-1}\mathbb{Z}^2 + \mathbf{B}_{i+1}^{-1}\boldsymbol{p}' = \mathbf{B}_{i+1}^{-1}\mathbb{Z}^2$ and on the right

$$\overset{\text{disjoint}}{\bigcup_{\boldsymbol{p} \in [\mathbb{Z}^2/\mathbf{R}_i^T\mathbb{Z}^2]}} \mathbf{B}_{i+1}^{-1}(\boldsymbol{p} + \boldsymbol{p}') + \mathbf{B}_i^{-1}\mathbb{Z}^2 = \overset{\text{disjoint}}{\bigcup_{\boldsymbol{p} \in [\mathbb{Z}^2/\mathbf{R}_i^T\mathbb{Z}^2]'}} \mathbf{B}_{i+1}^{-1}(\boldsymbol{p}) + \mathbf{B}_i^{-1}\mathbb{Z}^2,$$

where $\left[\mathbb{Z}^2/\mathbf{R}_i^T\mathbb{Z}^2\right]' \triangleq \left[\mathbb{Z}^2/\mathbf{R}_i^T\mathbb{Z}^2\right] + \boldsymbol{p}'$. The new partition of $\mathbf{B}_{i+1}^{-1}\mathbb{Z}^2$ on the right is the same as in (52), because the same sets appear. They are written differently because a different family $\left[\mathbb{Z}^2/\mathbf{R}_i^T\mathbb{Z}^2\right]$ of band designators is used. This is equivalent to choosing a different period as *the* period above.

In (52) passband centers for $H^{i+1}(\underline{\mathbf{k}} \cdot \mathbf{B}_{i+1})$ on the left are partitioned on the right into passband centers (the $\boldsymbol{p}_0$ term) and stopband centers (the other terms) of $H^i(\underline{\mathbf{k}} \cdot \mathbf{B}_i)$, so $H^i(\underline{\mathbf{k}} \cdot \mathbf{B}_i)$ suppresses $|\mathbf{R}_i| - 1$ of every $|\mathbf{R}_i|$ passbands of $H^{i+1}(\underline{\mathbf{k}} \cdot \mathbf{B}_{i+1})$ in the product $H^i(\underline{\mathbf{k}} \cdot \mathbf{B}_i)H^{i+1}(\underline{\mathbf{k}} \cdot \mathbf{B}_{i+1})$ that is steered in (51). Here is a hypothetical, nonsensical alternative to $H^i(\underline{\mathbf{k}} \cdot \mathbf{B}_i)$:

$$H_\Sigma^i(\underline{\mathbf{k}} \cdot \mathbf{B}_i) \triangleq \sum_{\boldsymbol{p} \in [\mathbb{Z}^2/\mathbf{R}_i^T\mathbb{Z}^2]} H^i\big((\underline{\mathbf{k}} - \mathbf{B}_{i+1}^{-1}\boldsymbol{p}) \cdot \mathbf{B}_i\big). \tag{53}$$

The $\underline{\mathbf{k}} = 0$ passband of $H^i(\underline{\mathbf{k}} \cdot \mathbf{B}_i)$ is shifted here to each stopband location in the period in turn (the shift to the passband location is a nonshift due to periodicity), so as an alternative it is nonsense because every band of interest in $H^i(\underline{\mathbf{k}} \cdot \mathbf{B}_i)$ becomes a passband in $H_\Sigma^i(\underline{\mathbf{k}} \cdot \mathbf{B}_i)$. These bands were co-located with the passbands of $H^{i+1}(\underline{\mathbf{k}} \cdot \mathbf{B}_{i+1})$, so we now have $H_\Sigma^i(\underline{\mathbf{k}} \cdot \mathbf{B}_i) \approx 1$ in the $H^{i+1}(\underline{\mathbf{k}} \cdot \mathbf{B}_{i+1})$ passbands. It seems that our masking-filter design goals have effortlessly come close to satisfying $H_\Sigma^i(\underline{\mathbf{k}} \cdot \mathbf{B}_i) = 1$, which would correspond

to meeting a more subtle constraint on $H^i(\underline{\mathbf{k}}\cdot\mathbf{B}_i)$ specified indirectly through (53). Two questions: Is it possible to meet it exactly? Would it be useful?

We'll answer both in coefficient space. In terms of $\mathbf{G}=\mathbf{B}_i^T\cdot\mathbf{B}_i$, the dot product $(\mathbf{B}_{i+1}^{-1}\boldsymbol{p})\cdot\mathbf{B}_i$ on the right in (53) becomes $(\mathbf{B}_i^{-1}\mathbf{R}_i^{-T}\boldsymbol{p})\cdot\mathbf{B}_i=(\mathbf{B}_i\mathbf{G}^{-1}\mathbf{R}_i^{-T}\boldsymbol{p})\cdot\mathbf{B}_i=\boldsymbol{p}^T\mathbf{R}_i^{-1}\mathbf{G}^{-1}\mathbf{B}_i^T\cdot\mathbf{B}_i=\boldsymbol{p}^T\mathbf{R}_i^{-1}$, so (53) holding for all $\underline{\mathbf{k}}\in\mathbb{R}^2$ is equivalent, by the linear independence of $\mathbf{b}_1$ and $\mathbf{b}_2$, to having

$$H^i_\Sigma(\boldsymbol{f})=\sum_{\boldsymbol{p}\in[\mathbb{Z}^2/\mathbf{R}_i^T\mathbb{Z}^2]}H^i(\boldsymbol{f}-\boldsymbol{p}^T\mathbf{R}_i^{-1}) \tag{54}$$

for all real two-element row vectors $\boldsymbol{f}$. Similarly, requiring $H^i_\Sigma(\underline{\mathbf{k}}\cdot\mathbf{B}_i)=1$ for all $\underline{\mathbf{k}}\in\mathbb{R}^2$ is equivalent to requiring $H^i_\Sigma(\boldsymbol{f})=1$. By (17) and (54),

$$H^i_\Sigma(\boldsymbol{f})=\sum_{\boldsymbol{p}\in[\mathbb{Z}^2/\mathbf{R}_i^T\mathbb{Z}^2]}\sum_{\boldsymbol{k}}h^i_{\boldsymbol{k}}\,\mathrm{e}^{-\mathrm{j}2\pi(\boldsymbol{f}-\boldsymbol{p}^T\mathbf{R}_i^{-1})\boldsymbol{k}}=\sum_{\boldsymbol{k}}h^i_{\boldsymbol{k}}\left(\sum_{\boldsymbol{p}\in[\mathbb{Z}^2/\mathbf{R}_i^T\mathbb{Z}^2]}\mathrm{e}^{\mathrm{j}2\pi\boldsymbol{p}^T\mathbf{R}_i^{-1}\boldsymbol{k}}\right)\mathrm{e}^{-\mathrm{j}2\pi\boldsymbol{f}\boldsymbol{k}}. \tag{55}$$

So $H^i_\Sigma(\boldsymbol{f})=1$ holds for all $\boldsymbol{f}$ if and only if for each k, either $h^i_{\boldsymbol{k}}=0$ or $z_{\boldsymbol{k}}=0$, where $z_{\boldsymbol{k}}$ is the quantity in large parentheses. When $\boldsymbol{k}\in\mathbf{R}_i\mathbb{Z}^2$, direct evaluation gives $z_{\boldsymbol{k}}=|\mathbf{R}_i|$, the number of elements in $\left[\mathbb{Z}^2/\mathbf{R}_i^T\mathbb{Z}^2\right]$. For other $\boldsymbol{k}$ we can use (54) and our "aside" above with $\boldsymbol{p}'\in\mathbb{Z}^2$ arbitrary and write

$$H^i_\Sigma(\boldsymbol{f}+\boldsymbol{p}'^T\mathbf{R}_i^{-1})=\sum_{\boldsymbol{p}\in[\mathbb{Z}^2/\mathbf{R}_i^T\mathbb{Z}^2]}H^i(\boldsymbol{f}-(\boldsymbol{p}-\boldsymbol{p}')^T\mathbf{R}_i^{-1})=\sum_{\boldsymbol{p}\in[\mathbb{Z}^2/\mathbf{R}_i^T\mathbb{Z}^2]-\boldsymbol{p}'}H^i(\boldsymbol{f}-\boldsymbol{p}^T\mathbf{R}_i^{-1})=H^i_\Sigma(\boldsymbol{f}).$$

This $H^i_\Sigma(\boldsymbol{f}+\boldsymbol{p}'^T\mathbf{R}_i^{-1})=H^i_\Sigma(\boldsymbol{f})$ can be written using (55) as

$$\sum_{\boldsymbol{k}}h^i_{\boldsymbol{k}}\left(z_{\boldsymbol{k}}\mathrm{e}^{-\mathrm{j}2\pi\boldsymbol{p}'^T\mathbf{R}_i^{-1}\boldsymbol{k}}\right)\mathrm{e}^{-\mathrm{j}2\pi\boldsymbol{f}\boldsymbol{k}}=\sum_{\boldsymbol{k}}h^i_{\boldsymbol{k}}\,z_{\boldsymbol{k}}\,\mathrm{e}^{-\mathrm{j}2\pi\boldsymbol{f}\boldsymbol{k}},$$

which holds for all $\boldsymbol{f}$ and $\boldsymbol{p}'$ and arbitrary h^i if and only if $z_{\boldsymbol{k}}\mathrm{e}^{-\mathrm{j}2\pi\boldsymbol{p}'^T\mathbf{R}_i^{-1}\boldsymbol{k}}=z_{\boldsymbol{k}}$ for all $\boldsymbol{k}\in\mathbb{Z}^2$ irrespective of the value of $\boldsymbol{p}'$, that is, if and only if at every $\boldsymbol{k}\in\mathbb{Z}^2$ either $z_{\boldsymbol{k}}=0$ or $\boldsymbol{p}'^T\mathbf{R}_i^{-1}\boldsymbol{k}$ for all $\boldsymbol{p}'\in\mathbb{Z}^2$. The latter is equivalent to $\boldsymbol{k}\in\mathbf{R}_i\mathbb{Z}^2$, for which we already know $z_{\boldsymbol{k}}=|\mathbf{R}_i|$, so $z_{\boldsymbol{k}}=0$ for $\boldsymbol{k}\notin\mathbf{R}_i\mathbb{Z}^2$.

We can now answer our two questions using Fourier pair $h^i_{\boldsymbol{k}}\,z_{\boldsymbol{k}}\leftrightarrow H^i_\Sigma(\boldsymbol{f})$ from (55). Yes, condition $H^i_\Sigma(\boldsymbol{f})=1$ can be met. It requires either

$$h^i_{\boldsymbol{k}}\,z_{\boldsymbol{k}}=\begin{cases}1 & \text{for } \boldsymbol{k}=0\\ 0 & \text{for } \boldsymbol{k}\neq 0\end{cases}\qquad\text{or}\qquad h^i_{\mathbf{R}_i\boldsymbol{n}}=\begin{cases}1/|\mathbf{R}_i| & \text{for } \boldsymbol{n}=0\\ 0 & \text{for other } \boldsymbol{n}\in\mathbb{Z}^2.\end{cases}$$

When optimizing we need only fix each $h^i_{\mathbf{R}_i\boldsymbol{n}}$ in advance. This is useful when the many zero coefficients mean hardware will not need to be built, and performance will seldom be significantly compromised. (To be sure, optimize with and without these extra conditions.)

The N in the "Nth-band" terminology is traditional and refers to the number of terms in the sum in (53), which constraint ties together the behavior of the "N" (here $|\mathbf{R}_i|$) bands of $H^i(\underline{\mathbf{k}}\cdot\mathbf{B}_i)$ of design interest, removing one degree of design freedom. This simplifies

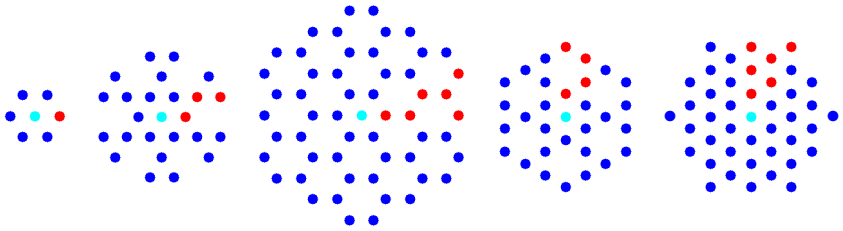

Figure 8: Coefficient locations for $h^1,\ldots,h^5$ of the Fig. 10 design example, each on its own scale. A few coefficients are optimized. Others are fixed or implied by symmetry.

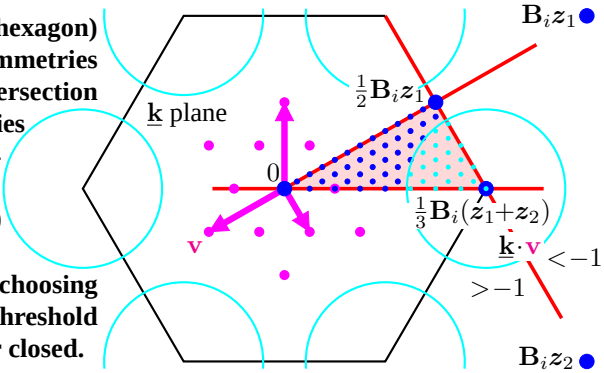

Figure 9: If Nth-band $H^i(\underline{\mathbf{k}}\cdot\mathbf{B}_i)$ has a period (hexagon) in the orbit, under the action of the common symmetries of passband lattices $\mathbf{B}_i\mathbb{Z}^2$ and $\mathbf{B}_{i+1}\mathbb{Z}^2$, of its intersection with a wedge (shaded), then forcing the symmetries on Nth-band h_i determines $H^i(\underline{\mathbf{k}}\cdot\mathbf{B}_i)$ on all passbands and stopbands (circles) from $H^i(\underline{\mathbf{k}}\cdot\mathbf{B}_i)$ on $\langle$wedge$\rangle\cap\langle$stopbands$\rangle$. On a wedge edge $\underline{\mathbf{k}}\cdot\mathbf{v}=0$ or -1 for a $\mathbf{v}\in(\frac{1}{6}\mathbf{B}_i)^{-1}\mathbb{Z}^2$ (center dots), so test constraint-grid $\underline{\mathbf{k}}\in\frac{1}{6\beta}\mathbf{B}_i\mathbb{Z}^2$ with $\underline{\mathbf{k}}\cdot\mathbf{v}\lesseqgtr 0$ or $-\beta$, choosing $\beta\in\mathbb{Z}$ large enough to constrain sidelobe peaks. Threshold changes of $\pm\frac{1}{2}$ cleanly make wedge edges open or closed.

optimization, because it is enough to constrain the stopband behavior of an Nth-band $H^i(\underline{\mathbf{k}}\cdot\mathbf{B}_i)$. The passband characteristic will follow automatically from (53).

We must take care when imposing symmetry on Nth-band array factors. Let $\mathcal{G}$ be a subgroup of the symmetry group of lattice basis $\mathbf{B}_i$, let $\mathbf{S}\in\mathcal{G}$, and suppose $\mathcal{L}$ is the symmetry of lattice $\mathbf{B}_i\mathbb{Z}^2$ with $\mathcal{L}\mathbf{B}_i=\mathbf{B}_i\mathbf{S}$. It could be that $\mathcal{L}$ is also a symmetry of sublattice $\mathbf{B}_{i+1}\mathbb{Z}^2=\mathbf{B}_i\mathbf{R}_i\mathbb{Z}^2$, in which case there is an invertible 2×2 integer matrix $\mathbf{S}'$ with $\mathbf{B}_i\mathbf{R}_i\mathbf{S}'=\mathcal{L}(\mathbf{B}_i\mathbf{R}_i)=(\mathcal{L}\mathbf{B}_i)\mathbf{R}_i=\mathbf{B}_i\mathbf{S}\mathbf{R}_i$ or, since $\mathbf{b}_1$ and $\mathbf{b}_2$ are linearly independent, $\mathbf{R}_i\mathbf{S}'=\mathbf{S}\mathbf{R}_i$. Suppose Nth-band array factor $H^i(\underline{\mathbf{k}}\cdot\mathbf{B}_i)$ is to be made symmetric with respect to $\mathbf{B}_i\mathcal{G}$ so that $0=h^i_{\mathbf{R}_i\boldsymbol{n}}=h^i_{\mathbf{S}\mathbf{R}_i\boldsymbol{n}}$ for $\boldsymbol{n}\neq 0$. If $\mathcal{L}$ is indeed a symmetry of sublattice $\mathbf{B}_{i+1}\mathbb{Z}^2=\mathbf{B}_i\mathbf{R}_i\mathbb{Z}^2$, then this requirement $0=h^i_{\mathbf{S}\mathbf{R}_i\boldsymbol{n}}=h^i_{\mathbf{R}_i\mathbf{S}'\boldsymbol{n}}$ is one the Nth-band condition requires anyway. If $\mathcal{L}$ is not a symmetry of sublattice $\mathbf{B}_i\mathbf{R}_i\mathbb{Z}^2$ however, then forcing $h^i_{\mathbf{S}\mathbf{R}_i\boldsymbol{n}}$ to zero is likely to adversely affect the optimized solution. We therefore impose symmetry of an Nth-band array factor $H^i(\underline{\mathbf{k}}\cdot\mathbf{B}_i)$ only with respect to $\mathbf{B}_i\mathcal{G}$ equal to the intersection of the symmetry groups of $\mathbf{B}_i\mathbb{Z}^2$ and sublattice $\mathbf{B}_{i+1}\mathbb{Z}^2=\mathbf{B}_i\mathbf{R}_i\mathbb{Z}^2$.

When optimizing an Nth-band array factor, optimization time and memory usage both benefit from constraining the array factor at only as many individual $\underline{\mathbf{k}}$ as necessary to adequately control its behavior. The strategy detailed in the Fig. 9 caption makes it only necessary to constrain $\frac{|\mathbf{R}_i|-1}{|\mathcal{G}|}$ stopbands, where $|\mathcal{G}|$ is the number of symmetries common to $\mathbf{B}_i\mathbb{Z}^2$ and $\mathbf{B}_{i+1}\mathbb{Z}^2$. When $\mathbf{B}=\mathbf{B}_\triangle$, this $|\mathcal{G}|$ takes values 12, 12, and 6 respectively for $|\mathbf{R}|_i$ values of 3, 4, and 7, so the respective number of stopbands that must be constrained is $\frac{1}{6}$, $\frac{1}{4}$, and one. The size of the optimization problem is also minimized by carefully choosing integer β, which controls the density of the stopband-constraint grid. Grid lattice $\frac{1}{6\beta}\mathbf{B}_i\mathbb{Z}^2$ has $(6\beta)^2$ points per period of passband lattice $\mathbf{B}_i\mathbb{Z}^2$, for $(6\beta)^2/|h^i|$ points per degree of coefficient freedom, a ratio for which a value of 10 to 20 is generally adequate. For the

latter then, set $\beta = \lceil \sqrt{20|h^i|}/6 \rceil$. It is also a good idea to space extra stopband-constraint points around the circular stopband edge, as this edge is where steep transitions begin.

The only non-Nth-band array factor in this paper is $H^5(\underline{\mathbf{k}} \cdot \mathbf{B}_5)$ of Fig. 10.

3.3. A Five-Layer Example Design

In this section we examine a high-performance beam-cluster system designed using lattices similar to $\mathbf{B}_\triangle \mathbb{Z}^2$ everywhere, a higher-performance system in the spirit of Fig. 4. We intentionally sought unreasonable performance—we didn't wish to design anyone's large array radar—by rolling off the main beam to the sidelobe floor very quickly. This yielded nice plots but a very large array, as array size is largely commensurate with the difficulty of the gymnastics performed by the beam edges to meet near-in sidelobe specifications.

We first set $\mathbf{B} = \mathbf{B}_1 = \alpha^{-1}\mathbf{B}_\triangle$ to space nearest neighbors of passband lattice $\mathbf{B}_1^{-1}\mathbb{Z}^2 = \alpha \mathbf{B}_\triangle^{-1}\mathbb{Z}^2$ by 2α times the visible-region radius. Parameter α should be as small as possible without permitting a second passband to enter into the visible region as a grating lobe when the beam is steered. Our plots—look ahead to Fig. 10—assume $\alpha = 1$, but changing α changes only the plotted radius of the visible-region circle. The images are unchanged.

In Fig. 4 digital steering vector $\underline{\mathbf{k}}_2$ is taken from some lattice that by construction is a superlattice of $\mathbf{B}_{N+1}^{-1}\mathbb{Z}^2$. This is not required but simply makes a diagram like Fig. 4 easier to draw. Similarly here and for no better reason, we took beam-steering lattice to be $\mathbf{B}_{N+1}^{-1}\mathbf{R}_s^{-T}\mathbb{Z}^2$. If Δk denotes its nearest-neighbor distance, a single beam step from zero then represents an angle of $\Delta\theta$ between $\mathbf{k}$ and boresight $\widehat{b}$ when

$$\sin \Delta\theta = \frac{\Delta k}{\text{visible-region radius}} = \frac{2\alpha\, \Delta k}{\text{passband spacing}} = 2\alpha \left(|\mathbf{R}_1| \cdots |\mathbf{R}_N| |\mathbf{R}_s| \right)^{-1/2} \quad (56)$$

so that $|\mathbf{R}_1| \cdots |\mathbf{R}_N| |\mathbf{R}_s| = \left(\frac{2\alpha}{\sin \Delta\theta} \right)^2 \approx 3300$ for $\Delta\theta = 2^\circ$ with $\alpha = 1$.

Subarray suppression of periodically repeating clusters is most efficient if, as in Fig. 4, the centerpoints of a cluster's beams, those $\mathbf{B}_{N+1}^{-1}\mathbf{R}_s^{-T}\mathbb{Z}^2$ points in some carefully chosen period of sublattice $\mathbf{B}_{N+1}^{-1}\mathbf{R}_s^{-T}\mathbf{R}_{\text{bigstep}}\mathbb{Z}^2$ used to step whole clusters, fit into as small a circular stopband as possible. When $|\mathbf{R}_{\text{bigstep}}| \in \{1+3n(n+1) : n = 0, 1, \ldots\} = \{7, 19, 37 \ldots\}$, the cluster's beam centerpoints can be placed in a multi-layer hexagonal pattern that is optimal for small $|\mathbf{R}_{\text{bigstep}}|$ as in the Fig. 10 example, where we arbitrarily chose $|\mathbf{R}_{\text{bigstep}}| = 19$.

Having $\mathbf{R}_s = \mathbf{R}_{\text{bigstep}}^T$ so that cluster steps come from $\mathbf{B}_{N+1}^{-1}\mathbb{Z}^2$ is also unnecessary but helpful for diagramming, so we required it for our own convenience. Here it implied $|\mathbf{R}_s| = 19$, so we needed $|\mathbf{R}_1| \cdots |\mathbf{R}_N| \approx 3300/19 \approx 173$. For efficiency a product of small determinants is best, and we could use only $|\mathbf{R}|$ values with entries in the $\mathbf{B}_\triangle$ row of the Fig. 7 table, so we simply factored all the integers near 173 to determine the options, of which only $192 = 4{\times}4{\times}4{\times}3$ limited the factors to 3's and 4's, efficient for Nth-band filters. Using 192 in place of 173 reduced beam-step size $\Delta\underline{\mathbf{k}}$ to the equivalent of 1.9°, and we had four options corresponding to the four ways to assign these factors of 192 to $|\mathbf{R}_1|, \ldots, |\mathbf{R}_4|$.

To evaluate the factoring options, we needed to know the stopband sizes, but these depended on which array factors would be steered to arrange beams within the cluster.

Table 1: To use Fig. 11, for each design option one needs the dB ratio $10\log_{10}\left(\frac{\textbf{area of } \mathcal{K}_s}{\textbf{area of period of } H^i(\underline{\mathbf{k}}\cdot\mathbf{B}_i)}\right)$ tabulated here for each Nth band subarray layer of the Fig. 10 design example.

		option (chosen)			
$\lvert\mathbf{R}_1\rvert$		3	4	4	4
$\lvert\mathbf{R}_2\rvert$		4	3	4	4
$\lvert\mathbf{R}_3\rvert$		4	4	3	4
$\lvert\mathbf{R}_4\rvert$		4	4	4	3
	$\mathcal{K}_s$ radius				
$H^1(\underline{\mathbf{k}}\cdot\mathbf{B}_1)$	$3\Delta k$	-20.5	-20.5	-20.5	-20.5
$H^2(\underline{\mathbf{k}}\cdot\mathbf{B}_2)$	$3\Delta k$	-15.7	-14.5	-14.5	-14.5
$H^3(\underline{\mathbf{k}}\cdot\mathbf{B}_3)$	$3\Delta k$	-9.7	-9.7	-8.4	-8.4
$H^4(\underline{\mathbf{k}}\cdot\mathbf{B}_4)$	Δk	-13.2	-13.2	-13.2	-12.0

Stepping whole clusters with $\mathbf{B}_{N+1}^{-1}\mathbb{Z}^2$ would make cluster size roughly $1/|\mathbf{R}_4|$ of the period of array factor $H^4(\underline{\mathbf{k}}\cdot\mathbf{B}_4)$. Imposing Nth-band requirement $H^i_\Sigma(\underline{\mathbf{k}}\cdot\mathbf{B}_i)=1$ would would leave no room for transition bands in the (53) sum, so an Nth-band $H^4(\underline{\mathbf{k}}\cdot\mathbf{B}_4)$ simply cannot suppress the entire cluster. Intra-cluster steering must therefore steer $H^4(\underline{\mathbf{k}}\cdot\mathbf{B}_4)$ as well as $H^5(\underline{\mathbf{k}}\cdot\mathbf{B}_5)$ so that that $H^4(\underline{\mathbf{k}}\cdot\mathbf{B}_4)$ can use stopbands only large enough to suppress a single beam. For any lattice similar to $\mathbf{B}_\triangle\mathbb{Z}^2$ the area of a period is $\frac{\sqrt{3}}{2}$ times the square of the nearest-neighbor distance, so using (56) and assuming $\mathcal{K}_s$ is a circle of radius $R\Delta k$,

$$\frac{\text{area of } \mathcal{K}_s}{\text{area of period of } H^i(\underline{\mathbf{k}}\cdot\mathbf{B}_i)} = \frac{\pi(R\Delta k)^2}{\frac{\sqrt{3}}{2}(\text{passband spacing})^2} = \frac{2\pi R^2/\sqrt{3}}{|\mathbf{R}_i|\cdots|\mathbf{R}_N||\mathbf{R}_s|}.$$

We somewhat arbitrarily set—these choices led to a large array—the radii of stopband-definition sets $\mathcal{K}_s$ (discussed just before Section 3.2) and then calculated the above ratio for each order-of-factors design option. The results are in Table 1. All columns begin with same dB value, and going down each increment is $10\log_{10}4\approx 6$ dB or $10\log_{10}3\approx 4.8$ dB.

The design aid in Fig. 11 for Nth-band spatial filters on lattices similar to $\mathbf{B}_\triangle\mathbb{Z}^2$ then allowed the required $|h_1|,\cdots,|h_4|$ to be estimated for each design option. We looked for a stopband suppression of around 40 dB—overdesign again. Ultimate array factor $H^5(\underline{\mathbf{k}}\cdot\mathbf{B}_5)$ is not Nth-band, so Fig. 11 does not apply and we instead estimated $|h^5|$ by trial-and-error optimization using a single-point passband set $\mathcal{K}_p$. The overall size of the array depends mostly on beamwidth and is substantially identical in all cases, so assuming h^1, h^2, and h^3 are analog while h^4 and h^5 are digital to ease intra-cluster steering, analog-subarraying cost can be compared using per-element cost metric

$$\frac{|h^1|}{|\mathbf{R}_1|} + \frac{|h^2|}{|\mathbf{R}_1|\,|\mathbf{R}_2|} + \frac{|h^3|}{|\mathbf{R}_1|\,|\mathbf{R}_2|\,|\mathbf{R}_3|}. \tag{57}$$

Each term is like the overlap γ used before but now adjusted for the spatial density of signals processed. Assignment of the factors of $|\mathbf{R}_1|\cdots|\mathbf{R}_N|$ to individual $|\mathbf{R}_i|$ properly also depends on the cost of steering the different designs, which we did not consider, and the different numbers of expensive digital receivers required at the input to the penultimate layer. No simple formula suffices; a full trade study is required for any real design. For this example, however, we settled on the design of Fig. 10.

The left image column in Fig. 10 contains the five array factors $H^i(\underline{\mathbf{k}}\cdot\mathbf{B}_i)$ on the same $\underline{\mathbf{k}}$ scale. For each of the last four, an image zoomed in to the square is shown immediately to the right, in the second column. Each of the first four third-column images shows, on the scale of the plot to its left, the maximum of the corresponding $G^i(\underline{\mathbf{k}}\cdot\mathbf{B}_i)$ (this $\underline{\mathbf{k}}$ is actually some $\underline{\mathbf{k}}+\Delta\underline{\mathbf{k}}$ for intra-cluster steering, but we suppress the clumsy notation here) taken over the 19 beams of the cluster. Overall array factor $G^1(\underline{\mathbf{k}}\cdot\mathbf{B}_1)$ appears again on the lower right at that row's zoom scale. Figure 8 shows the $h^1,\ldots,h^5$ coefficient locations on normalized scales. (Unnormalize the ith plot by scaling up by product $|\mathbf{R}_1|\cdots|\mathbf{R}_{i-1}|$.) Each is Nth band and represents coefficients invariant with respect to twelve symmetries.[2]

Table 2 details hardware requirements. The number of signals in s_{i-1} is $|g_i|$, so the array has $|g_1|=27,001$ elements and $|g_4|=295$ digital receivers. Ratio $\gamma_i \triangleq |h_i|/|\mathbf{R}_i|$ is the overlap of subarray layer i. Replacing $N{+}1$ in (47) with i makes (48) give a coefficient set, call it g_i', equivalent to subarray layers 1 through i as a subsystem, so the equivalent subarray feeding each A/D converter has $|g_3'|=1,837$ elements. Ratio $\gamma_i' \triangleq |g_i'|/(|\mathbf{R}_1|\cdots|\mathbf{R}_i|)$ is the overlap of these equivalent subarrays, of which those driving the A/D's have overlap $\sqrt{\gamma_3'}\approx 6.2$ in each dimension. This very high number corresponds to the extraordinary combined performance of array factors $H^1(\underline{\mathbf{k}}\cdot\mathbf{B}_1)$, $H^2(\underline{\mathbf{k}}\cdot\mathbf{B}_2)$, and $H^3(\underline{\mathbf{k}}\cdot\mathbf{B}_3)$ in Fig. 10. The total number of combiner inputs prior to the A/D's, the best overall measure of hardware complexity for a subarraying system, is $|h^1||g_2|+|h^2||g_3|+|h^3||g_4|=94,389$, about 3.5 per element. Compare to 9 for the poor-man's subarray example in Fig. 5.

The symmetry benefits are substantial also. Because of the Fig. 8 symmetries, this five-layer system can be realized with only 20 distinct weights relative to the center coefficients that set overall gain levels. Only 10 of those distinct weights are in the analog layers.

4. Summary and Conclusions

A general multi-layer technique for realizing array factors using spatial versions of the filter-decimate chains routine to DSP has been presented, and a significant design example has suggested that order-of-magnitude reductions in analog-subarraying hardware relative to the conventional subarray approach may be feasible. Noise performance has not been dealt with here, but its analysis should be straightforward and its results unremarkable.

It is important to realize that there is considerable room for improvement in array-factor design relative to what we have presented here, because the approach here was deliberately kept basic. The subarray array factors of the Fig. 10 example design were overdesigned as well as overspecified, because a small stopband was implemented in each $H^i(\underline{\mathbf{k}}\cdot\mathbf{B}_i)$ when it was a large stopband in $G^i(\underline{\mathbf{k}}\cdot\mathbf{B}_i)$ that was actually needed. A more-complex, "corrected" optimization formulation might lower hardware requirements further by eliminating overspecification of $H^i(\underline{\mathbf{k}}\cdot\mathbf{B}_i)$ and thereby permitting smaller $|h^i|$ values. Optimizing 1D filters for cascade performance is well developed [8], and the technique can be applied to array factors straightforwardly.

[2] We avoided $|\mathbf{R}_i|=7$ because seventh-band array factors on the triangular lattice are only symmetric with respect to six symmetries, which hurts performance. Adding $|\mathbf{R}_i|=7$ curves to Fig. 11 is not very useful.

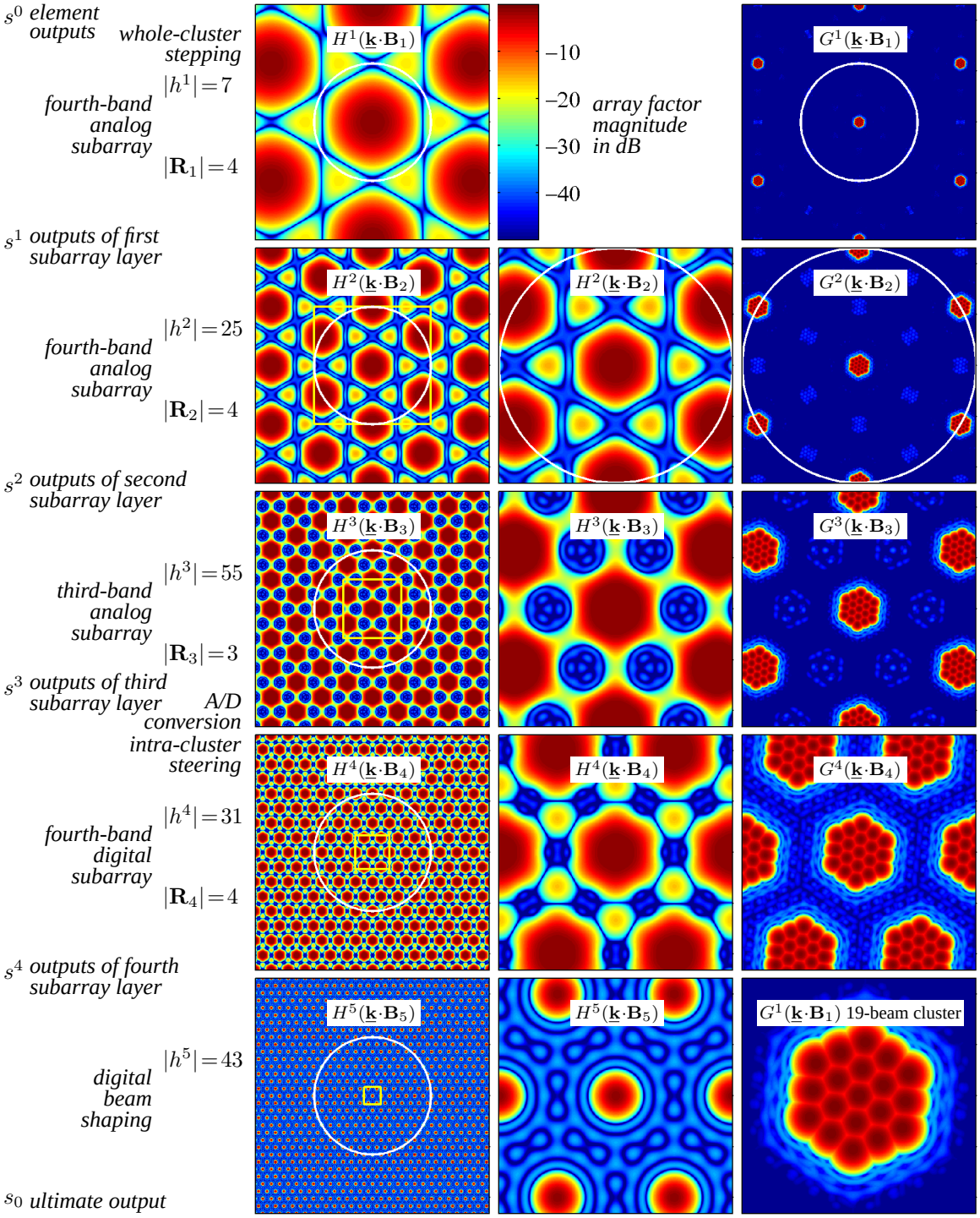

Figure 10: An intentionally overdesigned triangular array of 27,001 elements using three analog subarray layers, one digital subarray layer, and final digital beam shaping with 19-beam digital intra-cluster steering between the analog and digital layers. See Table 2 for key hardware numbers.

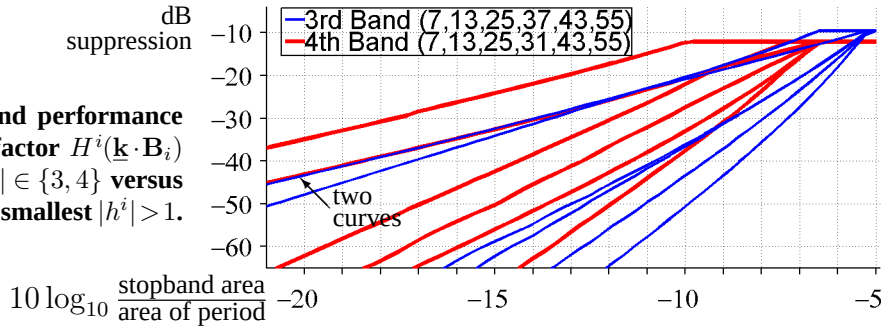

Figure 11: The stopband performance of an Nth-band array factor $H^i(\underline{\mathbf{k}}\cdot\mathbf{B}_i)$ when $\mathbf{B}_i = \mathbf{B}_\triangle$ and $|\mathbf{R}_i| \in \{3,4\}$ versus stopband size for the six smallest $|h^i| > 1$.

i	$\lvert h_i\rvert$	$\lvert g_i\rvert$	γ_i	$\lvert g_i'\rvert$	$\sqrt{\gamma_i'}$
1	7	27,001	1.75	7	1.3
2	25	6,427	6.25	133	2.9
3	55	1,327	18.33	1,837	6.2
4	31	295			
5	43	43			

number of elements
number of digital receivers
equivalent analog subarray size
overlap of equivalent analog subarrays in each direction

Table 2: Hardware requirements for the Fig. 10 example. Analog subarraying performance: equivalent overlap $\approx 6.2^2 \approx 38.3$. Analog subarraying cost: $(\,|h^1|\,|g_2| + |h^2|\,|g_3| + |h^3|\,|g_4|\,) \,/\, |g_1| \approx 3.5$. Cost/performance $\approx 3.5/38.3 \approx 9\%$. Compare: cost/performance $= 100\%$ for conventional subarrays.

References

[1] Y. Neuvo, C.-Y. Dong, and S. K. Mitra, "Interpolated Finite Impulse Response Filters," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 32, pp. 563–570, June 1984.

[2] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattice and Groups*, Springer-Verlag (New York), 2nd ed., 1988.

[3] I. N. Herstein, *Topics in Algebra*, Wiley (NY), 2nd ed., 1975.

[4] W. Rudin, *Real and Complex Analysis*, McGraw-Hill (New York), 3rd ed., 1987.

[5] S. Römisch, D. Popović, N. Shino, R. Lee, and Z. Popović, "Multi-beam discrete lens arrays with amplitude-controlled steering," in *Digest, 2003 IEEE MTT-S International Microwave Symp.*, no. 8–13 in vol. 3 (Tampa, http://ieeexplore.ieee.org/iel5/8599/27239/01210459.pdf), pp. 1669–1672, June 2003.

[6] D. P. Scholnik and J. O. Coleman, "Superdirectivity and SNR constraints in wideband array-pattern design," in *Proc. IEEE International Radar Conference* (Atlanta, http://ewh.ieee.org/soc/aes/Conferences.html), May 2001.

[7] J. O. Coleman, D. P. Scholnik, and P. E. Cahill, "Synthesis of a polarization-controlled pattern for a wideband array by solving a second-order cone program," in *Proc. IEEE Int'l Symp. on Antennas and Propagation* (Washington DC, http://apsursi2005.org/), July 3–8, 2005.

[8] J. O. Coleman, D. P. Scholnik, and J. J. Brandriss, "A specification language for the optimal design of exotic FIR filters with second-order cone programs," in *Proc. 36th IEEE Asilomar Conference on Signals, Systems, and Computers* (http://web.nps.navy.mil/∼asilomar), Nov. 2002.