



A comparison of vector symbolic architectures

Kenny Schlegel¹ · Peer Neubert¹ · Peter Protzel¹

Published online: 15 December 2021
© The Author(s) 2021

Abstract

Vector Symbolic Architectures combine a high-dimensional vector space with a set of carefully designed operators in order to perform symbolic computations with large numerical vectors. Major goals are the exploitation of their representational power and ability to deal with fuzziness and ambiguity. Over the past years, several VSA implementations have been proposed. The available implementations differ in the underlying vector space and the particular implementations of the VSA operators. This paper provides an overview of eleven available VSA implementations and discusses their commonalities and differences in the underlying vector space and operators. We create a taxonomy of available binding operations and show an important ramification for non self-inverse binding operations using an example from analogical reasoning. A main contribution is the experimental comparison of the available implementations in order to evaluate (1) the capacity of bundles, (2) the approximation quality of non-exact unbinding operations, (3) the influence of combining binding and bundling operations on the query answering performance, and (4) the performance on two example applications: visual place- and language-recognition. We expect this comparison and systematization to be relevant for development of VSAs, and to support the selection of an appropriate VSA for a particular task. The implementations are available.

Keywords Vector symbolic architectures · Hypervectors · High-dimensional computing · Hyperdimensional computing

1 Introduction

This paper is about selecting the appropriate Vector Symbolic Architecture (VSA) to approach a given task. But what is a VSA? VSAs are a class of approaches to solve computational problems using mathematical operations on large vectors. A VSA consists of a

✉ Kenny Schlegel
kenny.schlegel@etit.tu-chemnitz.de
Peer Neubert
peer.neubert@etit.tu-chemnitz.de
Peter Protzel
peter.protzel@etit.tu-chemnitz.de

¹ Faculty of Electrical Engineering, Chemnitz University of Technology, Chemnitz, Germany

particular vector space, for example $[-1, 1]^D$ with $D = 10,000$ (the space of 10,000-dimensional vectors with real numbers between -1 and 1) and a set of well chosen operations on these vectors. Although each vector from $[-1, 1]^D$ is primarily a subsymbolic entity without particular meaning, we can associate a symbolic meaning with this vector. To some initial atomic vectors, we can assign a meaning. For other vectors, the meaning will depend on the applied operations and operands. This is similar to how a symbol can be encoded in a binary pattern in a computer (e.g., encoding a number). In the computer, imperative algorithmic processing of this binary pattern is used to perform manipulation of the symbol (e.g., do calculations with numbers). The binary encodings in computers and operations on these bitstrings are optimized for maximum storage efficiency (i.e., to be able to distinguish 2^n different numbers in an n -dimensional bitstring) and for exact processing (i.e., there is no uncertainty in the encodings or the outcome of an operation). Vector Symbolic Architectures follow a considerably different approach:

- (1) Symbols are encoded in very large atomic vectors, much larger than would be required to just distinguish the symbols. VSAs use the additional space to introduce redundancy in the representations, usually combined with distributing information across many dimensions of the vector (e.g., there is no single bit that represents a particular property—hence a single error on this bit can not alter this property). As an important result, this redundant and distributed representation allows to also store compositional structures of multiple atomic vectors in a vector from the same space. Moreover, it is known from mathematics that in very high dimensional spaces *randomly* sampled vectors are very likely almost orthogonal (Kanerva 2009) (a result of the concentration of measure). This can be exploited in VSAs to encode symbols using *random* vectors and, nevertheless, there will be only a very low chance that two symbols are similar in terms of angular distance measures. Very importantly, measuring the angular distance between vectors allows us to evaluate a graded similarity relation between the corresponding symbols.
- (2) The operations in VSAs are mathematical operations that create, process and preserve the graded similarity of the representations in a systematic and useful way. For instance, an addition-like operator can overlay vectors and creates a representation that is similar to the overlaid vectors. Let us look at an example (borrowed from Kanerva (2009)): Suppose that we want to represent the country *USA* and its properties with symbolic entities—e.g., the currency *Dollar* and capital *Washington DC* (abbreviated *WDC*). In a VSA representation, each entity is a high-dimensional vector. For basic entities, for which we do not have additional information to systematically create them, we can use a *random* vector (e.g., sample from $[-1, 1]^D$). In our example, these might be *Dollar* and *WDC*—remember, these two high-dimensional random vectors will be very dissimilar. In contrast, the vector for *USA* shall reflect our knowledge that *USA* is related to *Dollar* and *WDC*. Using a VSA, a simple approach would be to create the vector for *USA* as a superposition of the vectors *Dollar* and *WDC* by using an operator $+$ that is called bundling: $R_{USA} = \textit{Dollar} + \textit{WDC}$. A VSA implements this operator such that it creates a vector R_{USA} (from the same vector space) that is *similar* to the input vectors—hence, R_{USA} will be similar to both *WDC* and *Dollar*.

VSAs provide more operators to represent more complex relations between vectors. For instance, a binding operator \otimes that can be used to create role-filler pairs and create and query more expressive terms like: $R_{USA} = \textit{Name} \otimes \textit{USA} + \textit{Curr} \otimes \textit{Dollar} + \textit{Cap} \otimes \textit{WDC}$,

with *Name*, *Curr*, and *Cap* being random vectors that encode these three roles. Why is this useful? We can now query for the currency of the USA by another mathematical operation (called unbinding) on the vectors and *calculate* the result by: $Dollar = R_{USA} \oslash Curr$. Most interestingly, this query would still work under significant amounts of fuzziness—either due to noise, ambiguities in the word meanings, or synonyms (e.g. querying with *monetary unit* instead of *currency*—provided that these synonym vectors are created in an appropriate way, i.e. they are similar to some extent). The following Sect. 2 will provide more details on these VSA operators.

Using embeddings in high-dimensional vector spaces to deal with ambiguities is well established in natural language processing (Widdows 2004). There, the objective is typically a particular similarity structure of the embeddings. VSAs make use of a larger set of operations on high-dimensional vectors and focus on the sequence of operations that generated a representation. A more exhaustive introduction to the properties of these operations can be found in the seminal paper of Kanerva (2009) and in the more recent paper (Neubert et al. 2019b). So far, they have been applied in various fields including medical diagnosis (Widdows and Cohen 2015), image feature aggregation (Neubert and Schubert 2021), semantic image retrieval (Neubert et al. 2021), robotics (Neubert et al. 2019b), to address catastrophic forgetting in deep neural networks (Cheung et al. 2019), fault detection (Kleyko et al. 2015), analogy mapping (Rachkovskij and Slipchenko 2012), reinforcement learning (Kleyko et al. 2015), long-short term memory (Danilhelka et al. 2016), pattern recognition (Kleyko et al. 2018), text classification (Joshi et al. 2017), synthesis of finite state automata (Osipov et al. 2017), and for creating hyperdimensional stack machines (Yerxa et al. 2018). Interestingly, also the intermediate and output layers of deep artificial neural networks can provide high-dimensional vector embeddings for symbolic processing with a VSA (Neubert et al. 2019b; Yilmaz 2015; Karunaratne et al. 2021). Although processing of vectors with thousands of dimensions is currently not very time efficient on standard CPUs, typically, VSA operations can be highly parallelized. In addition, there are also particularly efficient in-memory implementations of VSA operators possible (Karunaratne et al. 2020). Further, VSAs support distributed representations, which are exceptionally robust towards noise (Ahmad and Hawkins 2015), an omnipresent problem when dealing with real world data, e.g., in robotics (Thrun et al. 2005). In the long term, this robustness can also allow to use very power efficient stochastic devices (Rahimi et al. 2017) that are prone to bit errors but are very helpful for applications with limited resources (e.g., mobile computing, edge computing, robotics).

As stated initially, a VSA combines a vector space with a set of operations. However, based on the chosen vector space and the implementation of the operations, a different VSA is created. In the above list of VSA applications, a broad range of different VSAs has been used. They all use a similar set of operations, but the different underlying vector spaces and the different implementations of the operations have a large influence on the properties of each individual VSA. Basically, each application of a VSA raises the question: Which VSA is the best choice for the task at hand? This question gained relatively little attention in the literature. For instance, Widdows and Cohen (2015), Kleyko (2018), Rahimi et al. (2017) and Plate (1997) describe various possible vector spaces with corresponding bundling and binding operation but do not experimentally compare these VSAs on an application. A capacity experiment of different VSAs in combination with Recurrent Neuronal Network memory was done in Frady et al. (2018). However, the authors focus particularly on the application of the recurrent memory rather than the complete set of operators.

In this paper, we benchmark eleven VSA implementations from the literature. We provide an overview of their properties in the following Sect. 2. This section also presents a novel taxonomy of the different existing binding operators and discusses the algorithmic ramifications of their mathematical properties. A more practically relevant contribution is the experimental comparison of the available VSAs in Sect. 3 with respect to the following important questions: (1) *How efficiently can the different VSAs store (bundle) information into one representation?* (2) *What is the approximation quality of non exact unbind operators?* (3) *To what extend are binding and unbinding disturbed by bundled representations?* In Sect. 4, we complement this evaluation based on synthetic data with an experimental comparison on two practical applications that involve real-world data: the ability to encode context for visual place recognition on mobile robots and the ability to systematically construct symbolic representations for recognizing the language of a given text. The paper closes with a summary of the main insights in Sect. 5. Matlab implementations of all VSAs and the experiments are available online.¹

We want to emphasize the point that a detailed introduction to VSAs and their operators are beyond the scope of this paper—instead, we focus on a comparison of available implementations. For more basic introductions to the topic please refer to Kanerva (2009) or Neubert et al. (2019b).

2 VSAs and their properties

A VSA combines a vector space with a set of operations. The set of operations can vary but typically includes operators for bundling, binding, and unbinding, as well as a similarity measure. These operators are often complemented by a permutation operator which is important, e.g., to quote information (Gayler 1998). Despite their importance, since permutations work very similar for all VSAs, they are not part of this comparison. Instead we focus on differences between VSAs that can result from differences in one or multiple of the other components described in the following subsections. We selected the following implementations² (summarized in Table 1): the Multiply-Add-Permute (we use the acronyms **MAP-C**, **MAP-B** and **MAP-I**, to distinguish their three possible variations based on real, bipolar or integer vector spaces) from Gayler (1998), the Binary Spatter Code (**BSC**) from Kanerva (1996), the Binary Sparse Distributed Representation from Rachkovskij (2001) (**BSDC-CDT** and **BSDC-S** to distinguish the two different proposed binding operations), another Binary Sparse Distributed Representation from Laiho et al. (2015) (**BSDC-SEG**), the Holographic Reduced Representations (**HRR**) from Plate (1995) and its realization in the frequency domain (**FHRR**) from Plate (2003), Plate (1994), the Vector derived Binding (**VTB**) from Gosmann and Eliasmith (2019), which is also based on the ideas of Plate (1994), and finally an implementation called Matrix Binding of Additive Terms (MBAT) from Gallant and Okaywe (2013).

All these VSAs share the property of using high-dimensional representations (hyper-vectors). However, they differ in their specific vector spaces \mathbb{V} . Section 2.1 will introduce

¹ https://github.com/TUC-ProAut/VSA_Toolbox, additional supplemental material is also available at <https://www.tu-chemnitz.de/etit/proaut/vsa>.

² All VSAs are taken from the literature. However, in order to implement and experimentally evaluate them, we had to make additional design decisions for some. This led to the three versions of the MAP architecture from Gayler (1998).

Table 1 Summary of the compared VSAs

Name	Elements X of vector space \mathbb{V}	Initialization of an atomic vector x_i	Typical used sim. metric	Bundling	Binding		Unbinding		Ref.
					Commutative	Associative	Commutative	Associative	
MAP-C	$X \in \mathbb{R}^D$	$x_i \sim \mathcal{U}(-1, 1)$	Cosine sim.	Elem. addition with cutting	Elem. multipl.	Elem. multipl.	Elem. multipl.	Gayler (1998)	
MAP-I	$X \in \mathbb{Z}^D$	$x_i \sim \mathcal{B}(0.5) \cdot 2 - 1$	Cosine sim.	Elem. addition	Elem. multipl.	Elem. multipl.	Elem. multipl.	Gayler (1998)	
HRR	$X \in \mathbb{R}^D$	$x_i \sim \mathcal{N}(0, \frac{1}{D})$	Cosine sim.	Elem. addition with normalization	Circ. conv.	Circ. conv.	Circ. corr.	Plate (1995, 2003)	
VTB	$X \in \mathbb{R}^D$	$x_i \sim \mathcal{N}(0, \frac{1}{D})$	Cosine sim.	Elem. addition with normalization	VTB	VTB	transpose VTB	Gosmann and Ellis-smith (2019)	
MBAT	$X \in \mathbb{R}^D$	$x_i \sim \mathcal{N}(0, \frac{1}{D})$	Cosine sim.	Elem. addition with normalization	Matrix multipl.	Matrix multipl.	Inv. matrix multipl.	Gallant and Okaywe (2013)	
MAP-B	$X \in \{-1, 1\}^D$	$x_i \sim \mathcal{B}(0.5) \cdot 2 - 1$	Cosine sim.	Elem. addition with threshold	Elem. multipl.	Elem. multipl.	Elem. multipl.	Gayler and Levy (2009), Kleyko et al. (2018)	
BSC	$X \in \{0, 1\}^D$	$x_i \sim \mathcal{B}(0.5)$	Hamming dist.	Elem. addition with threshold	XOR	XOR	XOR	Kanerva (1996)	
BSDC-CDT	$X \in \{0, 1\}^D$	$x_i \sim \mathcal{B}(p \ll 1)$	Overlap	Disjunction	CDT	CDT	-	Rachkovskij (2001)	
BSDC-S	$X \in \{0, 1\}^D$	$x_i \sim \mathcal{B}(p \ll 1)$	Overlap	Disjunction (opt. thinning)	Shifting	Shifting	Shifting	Rachkovskij (2001)	
BSDC-SEG	$X \in \{0, 1\}^D$	$x_i \sim \mathcal{B}(p \ll 1)$	Overlap	Disjunction (opt. thinning)	Segment shifting	Segment shifting	Segment shifting	Laiho et al. (2015)	
FHRR	$X \in \mathbb{C}^D$	$x_i = e^{i\theta}$ $\theta \sim \mathcal{U}(-\pi, \pi)$	Angle distance	Angles of elem. addition	Elem. angle addition	Elem. angle addition	Elem. angle subtraction	Plate (1994)	

$\mathcal{U}(min, max)$ is the uniform distribution in range $[min, max]$. $\mathcal{N}(\mu, \sigma)$ defines the normal distribution with mean μ and variance σ . $\mathcal{B}(p)$ represents the Bernoulli distribution with probability p . D denotes the number of dimensions and p the density. The density p of BSDC architectures is $p \ll 1$. Rachkovskij (2001) showed that a probability of $p = \frac{1}{\sqrt{D}}$ results in the largest capacity. The density of BSDC-SEG corresponds to the number of segments (Laiho et al. 2015). See Sect. 2.1 for details. For each binding and unbinding operator the algebraic properties are listed (associative and commutative)—either check for true or a cross for false

properties of these high-dimensional vectors spaces and discuss the creation of hypervectors. The introduction emphasized the importance of a similarity measure to deal with the fuzziness of representations: instead of treating representations as same or different, VSAs typically evaluate their similarity. Section 2.2 will provide details of the used similarity metrics. Table 1 summarizes the properties of the compared VSAs. In order to solve computational problems or represent knowledge with a VSA, we need a set of operations: bundling will be the topic of Sect. 2.3 and binding and unbinding will be explained in Sect. 2.4. This section will also introduce a taxonomy that systematizes the significant differences in the available binding implementations. Finally, Sect. 2.5 will describe an example application of VSAs to analogical reasoning using the previously described operators. The application is similar to the USA-representation example from the introduction and will reveal important ramifications of non-self inverse binding operations.

2.1 Hypervectors: the elements of a VSA

A VSA works in a specific vector space \mathbb{V} with a defined set of operations. The generation of hypervectors from the particular vector space \mathbb{V} is an essential step in high-dimensional symbolic processing. There are basically three ways to create a vector in a VSA: (1) It can be the result of a VSA operation. (2) It can be the result of (engineered or learned) encoding of (real-world) data. (3) It can be an atomic entity (e.g. a vector that represents a role in a role-filler pair). For these role vectors, it is crucial that they are non-similar to all other unrelated vectors. Luckily, in the high-dimensional vectors spaces underlying VSAs, we can simply use random vectors since they are mutually quasi-orthogonal. From these three ways, the first will be the topic of the following subsections on the operators. The second way (encoding other data as vectors, e.g. by feeding an image through a ConvNet) is part of the Sect. 4.2 to encode images for visual place recognition. The third way of creating basic vectors is topic of this section since it plays an important role when using VSAs and varies significantly for the different available VSAs.

When selecting vectors to represent basic entities (e.g., symbols for which we do not know any relation that we can encode), the goal is to create maximally different encodings (to be able to robustly distinguish them in the presence of noise or other ambiguities). High-dimensional vector spaces offer plenty of space to push these vectors apart and moreover, they have the interesting property that *random* vectors are already very far away (Neubert et al. 2019b). In particular for angular distance measures, this means that two random vectors are very likely almost orthogonal (this is called quasi-orthogonal): If we sample the direction of vectors independent and identically distributed (i.i.d.) from a uniform distribution, the more dimensions the vectors have, the higher is the probability that the angle between two such random vectors is close to 90 degrees; for 10,000 dimensional real vectors, the probability to be in 90 ± 5 degrees is almost one. Please refer to Neubert et al. (2019b) for a more in-depth presentation and evaluation.

The quasi-orthogonality property is heavily used in VSA operations. Since the different available VSAs use different vector spaces and metrics (cf. Sect. 2.2), different approaches to create vectors are involved. The most common approach is based on real numbers in the continuous range. For instance, the Multiply-Add-Permute (**MAP-C**—C stands for continuous) architecture uses the real range of $[-1, 1]$. Other architectures such as **HRR**, **MBAT** as well as the **VTB** VSAs use a real range which is normally distributed with a mean of 0 and a variance of $1/D$ where D defines the number of dimensions. Another group uses binary vector spaces. For example, the Binary Spatter Code (**BSC**) and the binary MAP

(**MAP-B** as well as **MAP-I**) architecture generate the vectors in $\{0, 1\}$ or $\{-1, 1\}$. The creation of the binary values is based on a Bernoulli distribution with a probability of $p = 0.5$. By reducing the probability p , sparse vectors can be created for the **BSDC-CDT**, **BSDC-S** as well as the **BSDC-SEG** VSAs (where the acronym CDT means Context Depend Thinning, S means shifting, and SEG means segmentally shifting, all three are binding operations and are explained in Sect. 2.4). To initialize the **BSDC-SEG** correctly, we use the density p to calculate the number of segments $s = D \cdot p$ (this is needed for binding, as shown in Fig. 2) and randomly place a single 1 in each segment, all other entries are 0. The authors of Rachkovskij (2001) showed that a probability of $p = \frac{1}{\sqrt{D}}$ (D is the number of dimensions) achieves the highest capacity³ in the vector and is therefore used in these architectures. Finally, a complex vector space can be used. One example is the frequency Holographic Reduced Representations **FHRR** that uses complex numbers on the unit circle (the complex number in each vector dimension has length one) (Plate 1994). It is therefore sufficient to use uniformly distributed values in the range of $(-\pi, \pi]$ to define the angles of the complex values—thus, the complex vector can be stored using the real vector of angles θ . The complex numbers c can be computed from the angles θ by $c = e^{i \cdot \theta}$.

2.2 Similarity measurement

VSAs use similarity metrics to evaluate vector representations, in particular, to find relations between two given vectors (figure out whether the represented symbols have a related meaning). For example, given a noisy version of a hypervector as the output of a series of VSA operations, we might want to find the most similar elementary vector from a database of known symbols in order to decode this vector. A carefully chosen similarity metric is essential for finding the correct denoised vector from the database and to ensure a robust operation of VSAs. The term *curse of dimensionality* (Bellman 1961) describes the observation that algorithms that are designed for low dimensional spaces often fail in higher dimensional spaces—this includes similarity measures based on Euclidean distance (Beyer et al. 1999). Therefore, VSAs typically use other similarity metrics, usually based on angles between vectors or vector dimensions.

As shown in Table 1, the architectures **MAP-C**, **MAP-B**, **MAP-I**, **HRR**, **MBAT** and **VTB** use the cosine similarity (cosine of the angle) between vectors \mathbf{a} and $\mathbf{b} \in \mathbb{R}^D$: $s = \text{sim}(\mathbf{a}, \mathbf{b}) = \text{cos}(\mathbf{a}, \mathbf{b})$. The output is a scalar value ($\mathbb{R}^D \times \mathbb{R}^D \rightarrow \mathbb{R}$) within the range $[-1, 1]$. Note that -1 means collinear vectors in opposite directions and 1 means identical directions. A value of 0 indicates orthogonal vectors.

The binary vector space can be combined with different similarity metrics depending on the sparsity: Either the complementary Hamming Distance for binary dense vectors, like **BSC** or the overlap for binary sparse vectors as **BSDC-CDT**, **BSDC-S**, **BSDC-SEG** (the overlap can be normalized to the range $[0, 1]$ (0 means non-similar and 1 means similar)). Equation 1 shows the equation to compute the similarity (complementary and normalized Hamming Distance) between dense ($p = 0.5$) binary vectors (**BSC**) \mathbf{a} and $\mathbf{b} \in \{0, 1\}^D$, given the number of dimensions D .

³ The term capacity refers to the number of stored items in the auto-associative memory in Rachkovskij (2001)

$$s = \text{sim}(\mathbf{a}, \mathbf{b}) = 1 - \frac{\text{HammingDist}(\mathbf{a}, \mathbf{b})}{D} \quad (1)$$

The complex space needs yet another similarity measurement. As introduced in section 2.1, the complex architecture of Plate (1994) (**FHRR**) uses angles θ of complex numbers. To measure how similar two vectors are, the average angular distance is calculated (keep in mind, since the complex vectors have unit length, vectors \mathbf{a} and \mathbf{b} are from \mathbb{R}^D and only contain the angles θ):

$$s = \text{sim}(\mathbf{a}, \mathbf{b}) = \frac{1}{D} \cdot \sum_{i=1}^D \cos(a_i - b_i) \quad (2)$$

2.3 Bundling

VSA's use the bundling operator to superimpose (or overlay) given hypervectors (similar to what was done in the introductory example). Bundling aggregates a set of input vectors of space \mathbb{V} and creates an output vector of the same space \mathbb{V} that is similar to its inputs. Plate (Plate 1997) declared that the essential property of the bundling operator is the unstructured similarity preservation. It means: a bundle of vectors $\mathbf{A} + \mathbf{B}$ is still similar to vector \mathbf{A} , \mathbf{B} and also to another bundle $\mathbf{A} + \mathbf{C}$ that contains one of the input vectors. Since all compared VSA's implement bundling as an addition-like operator, the most commonly used symbol for the bundling operation is $+$.

The implementation is typically a simple element-wise addition. Depending on the vector space it is followed by a normalization step to the specific numerical range. For instance vectors of the **HRR**, **VTB** and **MBAT** have to be scaled to a vector length of one. Bundled vectors from the **MAP-C** are cut at -1 and 1 . The binary VSA's **BSC** and **MAP-B** use a threshold to convert the sums into the binary range of values. The threshold depends on the number of bundled vectors and is exactly half this number. Potential ties in case of an even number of bundled vectors are decided randomly. In the sparse distributed architectures, the logical OR function is used to implement the bundling operation. Since only a few values are non-zero, they carry most information and shall be preserved. For example, Rachkovskij (2001) do not apply thinning after bundling, however, in some application it is necessary to decrease the density of the bundled vector. For instance, the language recognition example in Sect. 4.1 requires a density constraint—we used a (empirically determined) maximum density of 50%. Besides the **BSDC** without thinning, the **MAP-I** does not need normalization as well—it accumulates the vectors within the integer range. The bundling operator in **FHRR** first converts the angle vectors to the form $e^{i\theta}$ before using element-wise addition. Afterward, the complex-valued vectors will be added. Then, only the angles of the resulting complex numbers are used and the magnitudes are discarded—the output are the new angles θ . The complete bundling step is shown in equation 3:

$$\mathbf{a} + \mathbf{b} = \text{angle}(e^{i\mathbf{a}} + e^{i\mathbf{b}}) \quad (3)$$

Due to its implementation in form of addition, bundling is commutative and associative in all compared VSA implementations except for the normalized bundling operations which are only approximately associative: $(\mathbf{A} + \mathbf{B}) + \mathbf{C} \approx \mathbf{A} + (\mathbf{B} + \mathbf{C})$.

2.4 Binding

The binding operator is used to connect two vectors, e.g., the role-filler pairs in the introduction. The output is again a vector from the same vector space. Typically, it is the most complex and most diverse operator of VSAs. Plate (Plate 1997) defines the properties of the binding as follows:

- the output is non-similar to the inputs: binding of **A** and **B** is non similar to **A** and **B**
- it preserves structured similarity: binding of **A** and **B** is similar to binding of **A'** and **B'**, if **A'** is similar to **A** and **B'** is similar to **B**
- an inverse of the operation exists (defined as unbinding with symbol \oslash)

The binding is typically indicated by the mathematical symbol \otimes .

Unbinding \oslash is required to recover the elemental vectors from the result of a binding (Plate 1997). Given a binding $\mathbf{C} = \mathbf{A} \otimes \mathbf{B}$, we can retrieve the elemental vectors **A** or **B** from **C** with the unbinding operator: $\mathbf{R} = \mathbf{A} \oslash \mathbf{C}$ (or $\mathbf{B} \oslash \mathbf{C}$). **R** is now similar to the vector **B** or **A** respectively.

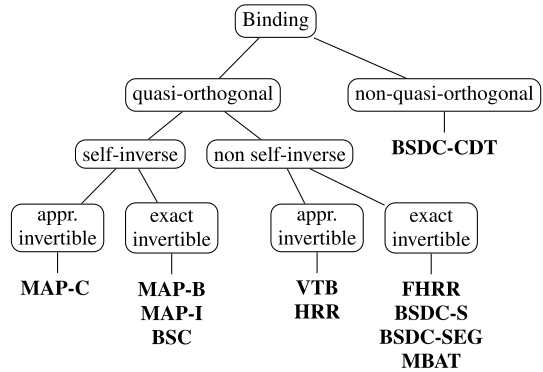
From a historical perspective, one of the first ideas to associate connectionist representations goes back to Smolensky (1990). He uses the tensor product (the outer product of given vectors) to compute a representation that combines all information of the inputs. To recover (unbind) the input information from the created matrix, it requires only the normalized inner product of the vector with the matrix (the tensor product). Based on this procedure, it is possible to perform exact binding and unbinding (recovering). However, using the tensor product creates a problem: the output of the tensor product of two vectors is a matrix and the size of the representation grows with each level of computation. Therefore, it is preferable to have binding operations (and corresponding unbinding operations) that *approximate* the result of the outer product in a vector ($\mathbb{V} \times \mathbb{V} \rightarrow \mathbb{V}$). Thus, according to Gayler (2003) a VSA's binding operation is basically a tensor product representation followed by a function to preserve the dimensionality of the input vectors. For instance, Frady et al. (2021) shows that the Hadamard product in the MAP VSA is a function of the outer product. Based on this dimensionality preserving definition, several binding and unbinding operations have been developed specifically for each vector domain. These different binding operations can be arranged in the taxonomy shown in Fig. 1.

The existing binding implementations can be basically divided into two types: quasi-orthogonal and non-quasi-orthogonal (see Fig. 1). Quasi-orthogonal bindings explicitly follow the properties of Plate (Plate 1997) and generate an output that is dissimilar to their inputs. In contrast, the output of a non-quasi-orthogonal binding will be similar to the input. Such a binding operation requires additional computational steps to achieve the properties specified by Plate (for example a nearest-neighbor search in an item memory (Rachkovskij 2001)).

On the next level of the taxonomy, quasi-orthogonal bindings can be further distinguished into self-inverse and non self-inverse binding operations. Self-inverse refers to the property that the inverse of the binding is the binding operation itself (unbinding = binding)⁴. The opposite is the non self-inverse binding: it requires an additional unbinding operator (inverse of the binding). Finally, each of these nodes can be separated into approximate

⁴ It should be noted that the operator is commonly referred to as self-inverse, but it is rather the vector that has this property and not the operator.

Fig. 1 Taxonomy of different binding operations. The VSAs that use each binding are printed in bold (see the Table 1 for more details)



and exact invertible binding (unbinding). For instance, the Smolensky tensor product is an exact invertible binding, because the unbinding produces exactly the same vector as in the input of the binding: $\mathbf{a} \otimes (\mathbf{a} \otimes \mathbf{b}) = \mathbf{b}$. The approximate inverse produces an unbinding output which is similar to the input of the binding, but not the same: $\mathbf{a} \otimes (\mathbf{a} \otimes \mathbf{b}) \approx \mathbf{b}$.

An quasi-orthogonal binding can be, for example, implemented by *element-wise multiplication* (as in Gayler (1998)). In case of bipolar values (± 1), element-wise multiplication is self-inverse, since $1^2 = -1^2 = 1$. The self-inverse property is essential for some VSA algorithms in the field of analogical reasoning (this will be the topic of Sect. 2.5). Element-wise multiplication is, for example, used in the **MAP-C**, **MAP-B** and **MAP-I** architectures. An important difference is that for the continuous space of **MAP-C** the unbinding is only approximate while it is exact for the binary space in **MAP-B**. For **MAP-I** it is exact for elementary vectors (from $\{-1, 1\}$) and approximate for processed vectors. Compared to the Smolensky tensor product, element-wise multiplication approximates the outer product matrix by its diagonal. Further, the element-wise multiplication is both commutative and associative (cf. Table 1).

Another self-inverse binding with an exact inverse is defined in the **BSC** architecture. It uses the *exclusive or (XOR)* and is equivalent to the element-wise multiplication in the bipolar space. As expected, the XOR is used for both binding and unbinding – it provides an exact inverse. Additionally, it is commutative and associative like element-wise multiplication.

The second category within the quasi-orthogonal bindings in our taxonomy in Fig. 1 are non self-inverse bindings. Two VSAs have an approximate unbinding operator. Binding of the real-valued vectors of the **VTB** architecture are computed using *Vector Derived Transformation (VTB)* as described in Gosmann and Eliasmith (2019). They use a matrix multiplication for binding and unbinding. The matrix is constructed from the second input vector \mathbf{b} , and multiplied with the first vector \mathbf{a} afterward. Equation 4 formulates the VTB as binding where V'_b represents a square matrix (Eq. 5) which is the reshaped vector b .

$$\mathbf{c} = \mathbf{a} \otimes \mathbf{b} = V'_b \cdot \mathbf{a} = \begin{bmatrix} V'_b & 0 & 0 \\ 0 & V'_b & 0 \\ 0 & 0 & \ddots \end{bmatrix} \mathbf{a} \tag{4}$$

$$V'_b = d^{\frac{1}{4}} \begin{bmatrix} b_1 & b_2 & \dots & b_{d'} \\ b_{d'+1} & b_{d'+2} & \dots & b_{2d'} \\ \vdots & \vdots & \ddots & \vdots \\ b_{d-d'+1} & b_{d-d'+2} & \dots & b_d \end{bmatrix}, d' = \sqrt{D} \tag{5}$$

This specifically designed transformation matrix (based on the second vector) provides a stringent transformation of the first vector which is invertible (i.e. it allows unbinding). This unbinding operator is identical to binding in terms of matrix multiplication, but the transposed matrix V_b is used for calculation, as shown in the Eq. 6. These binding and bundling operations are neither commutative nor associative.

$$\mathbf{a} \approx \mathbf{b} \oslash \mathbf{c} = V_b^T \mathbf{c} \tag{6}$$

Another approximated non self-invertible binding is part of the **HRR** architecture: the *circular convolution*. Binding of two vectors \mathbf{a} and $\mathbf{b} \in \mathbb{R}^D$ with circular convolution is calculated by:

$$\mathbf{c} = \mathbf{a} \otimes \mathbf{b} : c_j = \sum_{k=0}^{D-1} b_k a_{\text{mod}(j-k,D)} \text{ with } j \in \{0, \dots, D-1\} \tag{7}$$

Circular convolution approximates Smolensky’s outer product matrix by sums over all of its (wrap-around) diagonals. For more details please refer to Plate (1995). Based on the algebraic properties of convolution, this operator is commutative as well as associative. However, convolution is not self-inverse and requires a specific unbinding operator. The circular correlation (Eq. 8) provides an approximated inverse of the circular convolution and is used for unbinding. It is neither commutative nor associative.

$$\mathbf{a} \approx \mathbf{b} \oslash \mathbf{c} : a_j = \sum_{k=0}^{D-1} b_k c_{\text{mod}(k+j,D)} \text{ with } j \in \{0, \dots, D-1\} \tag{8}$$

A useful property of the convolution is that it becomes an element-wise multiplication in the frequency domain (complex space). Thus, it is possible to operate entirely in the complex vector space and use the element-wise multiplication as the binding operator (Plate 1994). This leads to the **FHRR** VSA with an exact invertible and non self-inverse binding⁵ as shown in the taxonomy in Fig. 1. With the constraints described in Sect. 2.1 (using complex values with a length of one), the computation of binding and unbinding becomes more efficient. Given two complex numbers c_1 and c_2 with angles θ_1 and θ_2 and length 1, multiplication of the complex numbers becomes an *addition* of the angles:

$$c_1 \cdot c_2 = e^{i \cdot \theta_1} \cdot e^{i \cdot \theta_2} = e^{i \cdot (\theta_1 + \theta_2)} \tag{9}$$

The same procedure applies to unbinding but with the angles of the conjugates of one of the given vectors—hence, it is just a *subtraction* of the angles θ_1 and θ_2 . Note that a modulo operation with 2π (angles on the complex plane are in the range of $(-\pi, \pi)$) must follow the addition or subtraction. Based on this assumption, it is possible to operate only with

⁵ It should be noted that there are relations between operations of different VSAs and between self-inverse and non self-inverse bindings: If the angles of an FHRR are quantized to two levels (e.g., $\{0, \pi\}$), the binding becomes self-inverse and equivalent to binary VSAs like BSC or MAP-B.

the angles rather than the whole complex numbers. Since the addition is associative and commutative, the binding is as well. But analog to the unbinding operation, subtraction is non-commutative and non-associative—therefore is also the unbinding. At this point we would like to emphasize that **HRR** and **FHRR** are basically functionally equivalent – the operations are performed either in spatial or frequency domain. However, the assumption of unit magnitudes in **FHRR** distinguishes both and simplifies the implementation of the binding. Moreover, in contrast to **FHRR**, **HRR** uses an approximate unbinding because it is more stable and robust against noise compared to an exact inverse (Plate 1994, p. 102).

In the following, we describe the two sparse VSAs with an quasi-orthogonal, exact invertible and non self-inverse binding: the **BSDC-S** (binary sparse distributed representations with shifting) and the **BSDC-SEG** (sparse vectors with segmental shifting as in Laiho et al. (2015)). The shifting operation allows to encode hypervectors into a new representation which is dissimilar to the input. Either the entire vector is shifted by a certain number or divided into segments and each segment is shifted individually by different values. The former goes as follows: Given two vectors, the first will be converted to a single hash-value (e.g. use the on-bits' position indices). Afterwards, the second vector is shifted by this hash-value (circular shifting). This operation has an exact inverse (shifting in the opposite direction), but it is neither commutative nor associative.

The latter (segment-wise shifting—**BSDC-SEG**) includes additional computing steps: As described in Laiho et al. (2015), the vectors are split into segments of the same length. Preferably, the number of segments depends on the density and is equal to the number of on-bits in the vector—thus, we have one on-bit per segment in average. For better understanding, see Fig. 2 for binding vector a with vector b . Each of those vectors has m segments (gray shaded boxes) with n values (bits). The position of the first on-bit in each segment of the vector gives one index per segment. Next, the segments of the second vector b will be circularly shifted by these indices (see the resulting vector in the figure). Like the **BSDC-S**, the unbinding is just a simple shifting by the negated indices of the vector a . Since the binding of this VSA resembles an addition of the segment indices, it is both commutative and associative. In contrast, the unbinding operation is a subtraction of the indices of vector a and b and is neither commutative nor associative. As mentioned earlier, different binding operations can be related. As another example, the binding operation of **BSDC-SEG** corresponds to an angular representation as in **FHRR** with m elements quantized to n levels.

The last VSA with an exact invertible binding mechanism is **MBAT**. It is similar to the earlier mentioned **VTB** binding that constructs a matrix to bind two vectors. **MBAT** (Gallant and Okaywe 2013) uses matrices with a size of $D \times D$ to bind vectors of length D —this procedure is similar to the Smolenskys tensor product. The binding matrix must be orthonormal and can be transposed to unbind a vector. To avoid creating a completely new matrix for each binding, Tissera and McDonnell (2014) uses an initial orthonormal matrix M and manipulates it for each binding. It uses the exponentiation of the initial matrix M by an arbitrary index i , resulting in a matrix M_i that is still orthonormal but after binding gives a different result than the initial matrix M . For our experimental comparison, we randomly sampled the initial matrix from a uniform distribution and convert it to an orthonormal matrix with the singular value decomposition. Since exponentiation of the initial matrix M leads to a high computational effort, we approximate the matrix manipulation by shifting the rows and the columns by the appropriate index of the role vector. This index is calculated with a hash-value of the role vector (simple summation over all indices of elements greater than zero). However, like the **VTB** VSA, the **MBAT** binding and unbinding are neither commutative nor associative.

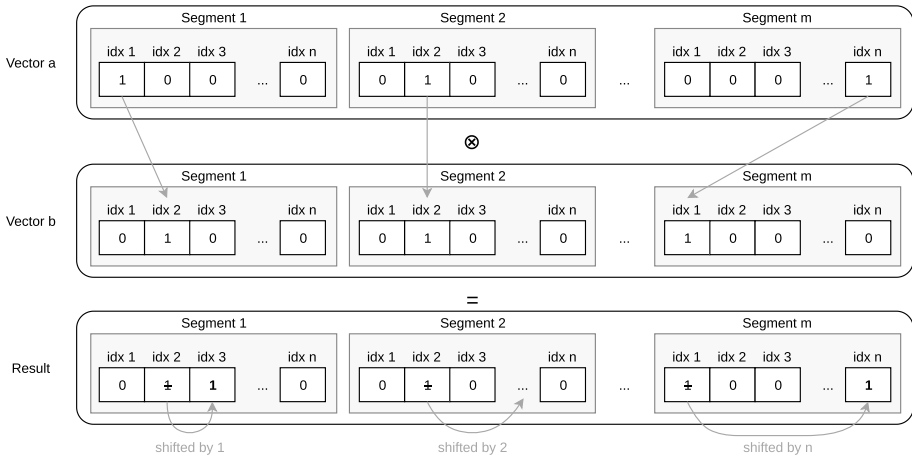


Fig. 2 Segment-wise shifting for binding sparse binary vectors a and b

According to Fig. 1, there is one VSA that uses a non-orthogonal binding. The **BSDC-CDT** from Rachkovskij (2001) introduces a binding operator for sparse binary vectors with an additive operator: the disjunction (logical OR). Since disjunction of sparse vectors can produce up to twice the number of on bits, they propose a *Context Depend Thinning (CDT)* procedure to thin vectors after the disjunction. The complete CDT procedure is described in Rachkovskij and Kussul (2001). Since this binding operation creates an output that is similar to the inputs, it is in contrast to Plate’s (1997) properties of binding operators (from the beginning of this section). As a consequence, instead of using unbinding to retrieve elemental vectors, the similarity to all elemental vectors has to be used to find the most similar ones. In contrast to the previously discussed quasi-orthogonal binding operations, here, additional computational steps are required to achieve the properties of the binding procedure defined by Plate (1997). Particularly, if the CDT is used for consecutive binding and bundling (e.g., bundling role-filler pairs can be seen as two levels—first is binding and second is bundling), this requires to store the specific level (binding at first level and bundling at the second level). During retrieval, the similarity search (unbinding) must be done in the corresponding level of binding, because this binding operator preserves the similarity of all bound vectors (in this example, every elemental vector is similar to the final representation after binding and bundling). Based on such iterative search (from level to level), the CDT binding needs more computational steps and is not directly comparable with the other binding operations. Therefore, the later experimental evaluations will use the segment-wise shifting as binding and unbinding for both the BSDC-S and BSDC-SEG VSAs instead of the CDT.

Finally, we want to emphasize the different complexities of the binding operations. Based on a comparison in Kelly et al. (2013), for D dimensional vectors, the complexities (number of computing steps) of binding two vectors are as follows:

- element-wise multiplication
(MAP-C, MAP-B, BSC, FHRR): $O(D)$
- circular conv. (HRR): $O(D \log D)$
- matrix binding (MBAT, VTB): $O(D^2)$

- sparse shifting (BSDC-S, BSDC-SEG)⁶: $O(D)$

2.5 Ramifications of non self-inverse binding

Section 2.4 distinguished two different types of binding operations: self-inverse and non self-inverse. We want to demonstrate possible ramifications of this property using the classical example from Pentti Kanerva on analogical reasoning (Kanerva 2010): “What is the Dollar of Mexico?” The task is as follows: Similar to the representation of the country *USA* ($R_{USA} = Name \otimes USA + Curr \otimes Dollar + Cap \otimes WDC$) from the example in the introduction, we can define a second representation of the country *Mexico*:

$$R_{Mex} = Name \otimes Mex + Curr \otimes Peso + Cap \otimes MXC \quad (10)$$

Given these two representations, we, as humans, can answer Kanerva’s question by analogical reasoning: Dollar is the currency of the USA, the currency of Mexico is Peso, thus the answer to the above question is “Peso”. This procedure can be elegantly implemented using a VSA. However, the method described in Kanerva (2010) only works with self-inverse bindings, such as **BSC** and **MAP**. To understand why, we will explain the VSA approach more in detail: Given are the records of both countries R_{Mex} and R_{USA} (the latter is written out in the introduction). In order to evaluate analogies between these two countries, we can combine all the information from these two representations into a single vector using binding. This creates a mapping F :

$$F = R_{USA} \otimes R_{Mex} \quad (11)$$

With the resulting vector representation we can answer the initial question (“What is the Dollar of Mexico?”) by binding the query vector (Dollar) to the mapping:

$$A = Dol \otimes F \approx Peso \quad (12)$$

The following explains why this actually works. Equation 11 can be examined based on the algebraic properties of the binding and bundling operations (e.g. binding distributes over bundling). In case of a self-inverse binding (cf. taxonomy in Fig. 1), the following terms result from Eq. 11 (we refer to Kanerva (2010) for a more detailed explanation):

$$F = (USA \otimes Mex) + (Dol \otimes Peso) + (WDC \otimes MXC) + N \quad (13)$$

Based on the self-inverse property, terms like $Curr \otimes Curr$ cancel out (i.e. they create a ones-vector). Since binding creates an output that is not similar to the inputs, other terms, like $Name \otimes Curr$, can be treated as noise and they are summarized in the term N . The noise terms are dissimilar to all known vectors and basically behave like random vectors (which are quasi-orthogonal in high-dimensional spaces). Binding the vector Dol to the mapping F of USA and Mexico (Eq. 12) creates vector A in Eq. 14 (only the most important terms are shown). The part $Dol \otimes (Dol \otimes Peso)$ is important because it reduces to $Peso$, again, based on the self-inverse property. As before, the remaining terms behave like noise that is bundled with the representation of $Peso$. Since the elemental vectors (representations for, e.g., *Dollar* or *Peso*) are randomly generated, they are highly robust against

⁶ Number of computational steps also depends on the density p .

noise. That is why the resulting vector A is still very similar to the elemental vector for $Peso$.

$$A = Dol \otimes ((USA \otimes Mex) + (Dol \otimes Peso) + \dots + N) \tag{14}$$

Notice, the previous description is only a brief summary to the ‘‘Dollar of Mexico’’ example. We refer to Kanerva (2010) for more details.

However, we can see that the computation is based on a self-inverse binding operation. As described in Sect. 2 and the taxonomy in Fig. 1, some VSAs have no self-inverse binding and need an unbind operator to retrieve elemental vectors.

The above described approach (Kanerva 2010) has the particularly elegant property that all information about the two records is stored in the single vector F and once this vector is computed, any number of queries can be done, each with a single operation (Eq. 12). However, if we relax this requirement, we can address the same task with the two-step approach described in Kanerva et al. (2001, p. 265). This also relaxes the requirement of a self-inverse binding and uses unbinding instead:

$$A = R_{Mex} \oslash (R_{USA} \oslash Dol) \tag{15}$$

After simplification to the necessary terms (all other terms are represented as noise N), we get equation 16.

$$\begin{aligned} A &= \underbrace{(Curr)}_{Role} \otimes \underbrace{(Peso)}_{Filler} \oslash \left(\underbrace{(Curr)}_{Role} \otimes \underbrace{(Dol)}_{Filler} \right) \oslash \underbrace{(Dol)}_{Filler} + N \\ A &= \underbrace{(Curr)}_{Role} \otimes \underbrace{(Peso)}_{Filler} \oslash \underbrace{(Curr)}_{Role} + N \\ A &= Peso + N \end{aligned} \tag{16}$$

It can be seen that it is in principle possible to solve the task ‘What is the dollar of Mexico?’ with non-self-inverse binding operators. However, this requires storing more vectors (both R_{Mex} and R_{USA} are stored) and additional computational effort.

In the same direction, Plate (1995) emphasized the need for a ‘readout’ machine for the HRR VSA to decode chunked sequences (hierarchical binding). It retrieves the trace iteratively and finally generates the result. Transferred to the given example: first, we have to figure out the meaning of *Dollar* (it is the currency of the USA) and query the result (*Currency*) on the representation of Mexico afterward (resulting in *Peso*). Such a readout requires more computation steps caused by iteratively traversing of the hierarchy tree (please see (Plate 1995) for more details). Presumably, this is a general problem of all non self-inverse binding operations.

3 Experimental comparison

After the discussion of theoretical aspects in the previous section, this section provides an experimental comparison of the different VSA implementations using three experiments. The first evaluates the bundling operations to answer the question *How efficiently can the different VSAs store (bundle) information into one representation?* The topic of the second experiment are the binding and unbinding operations. As described in Sect. 2.4 and the taxonomy in Fig. 1, some binding operations have an approximate inverse. Hence, the

second experiment evaluates the question *How good is the approximation of the binding inverse?* Finally, the third experiment focuses on the combination of bundling and binding and the ability to recover noisy representations. There, the leading question is: *To what extent are binding and unbinding disturbed by bundled representations?*

A note on the evaluation setup We will base our evaluation on the required *number of dimensions* of a VSA to achieve a certain performance instead of the physical *memory consumption* or *computational effort* - although the storage size and the computational effort per dimension can vary significantly (e.g. between a binary vector and a float vector). The main reason is that the actual resource demands of a single VSA might vary significantly dependent on the capabilities and limitations of the underlying hard- and software, as well as the current task. For example, it is well-known that HRR representations do not require a high precision for many tasks (Plate 1994, p. 67). However, low resolution data types (e.g. half-precision floats or less) might not be available in the used programming language. Instead, using the number of dimensions introduces a bias towards VSAs with high memory requirements per dimension, however, the values are supposed to be simple to convert to actual demands given a particular application setup.

3.1 Bundling capacity

We evaluate the question *How efficiently can the different VSAs store (bundle) information into one representation?* We use an experimental setup similar to Neubert et al. (2019b), extend it with varying dataset sizes and varying numbers of dimensions, and use it to experimentally compare the eleven VSAs. For each VSA, we create a database of $N = 1,000$ random elementary vectors from the underlying vector space \mathbb{V} . It represents basic entities stored in a so-called item memory. To evaluate the bundle capacity of this VSA, we randomly chose k elementary vectors (without replacement) from this database and create their superposition $B \in \mathbb{V}$ using the VSA's bundle operator. Now the question is whether this combined vector B is still similar to the bundled elementary vectors. To answer this question, we query the database with the vector B to obtain the k elementary vectors, which are the most similar to the bundle B (using the VSA's similarity metric). The evaluation criterion is the accuracy of the query result: the ratio of correctly retrieved elementary vectors on the k returned vectors from the database.⁷

The capacity depends on the dimensionality of \mathbb{V} . Therefore we range the number of dimensions D in $4 \dots 1156$ (since **VTB** needs even roots the number of dimensions is computed by i^2 with $i = 2 \dots 34$) and evaluate for k in $2 \dots 50$. We use $N = 1,000$ elementary vectors. To account for randomness, we repeat each experiment 10 times and report means.

Figure 3 shows the results of the experiment in form of a heat-map for each VSA, which encodes the accuracies of all combinations of number of bundled vectors and number of dimensions in colors. The warmer the color, the higher the achieved accuracy with a particular number of dimensions to store and retrieve a certain number of bundled vectors. One important observation is the large dark red areas (close to perfect accuracies) achieved by the **FHRR** and **BSDC** architectures. Also remarkable is the fast transition from very low accuracy (blue) to perfect accuracy (dark red) for the **BSDC** architectures; dependent on the number of dimensions, bundling will either fail or work almost perfectly. Presumably,

⁷ This experimental setup is closely related to Bloom filters that can efficiently evaluate whether an element is part of a set. Their relation to VSAs is discussed in Kleyko et al. (2020).

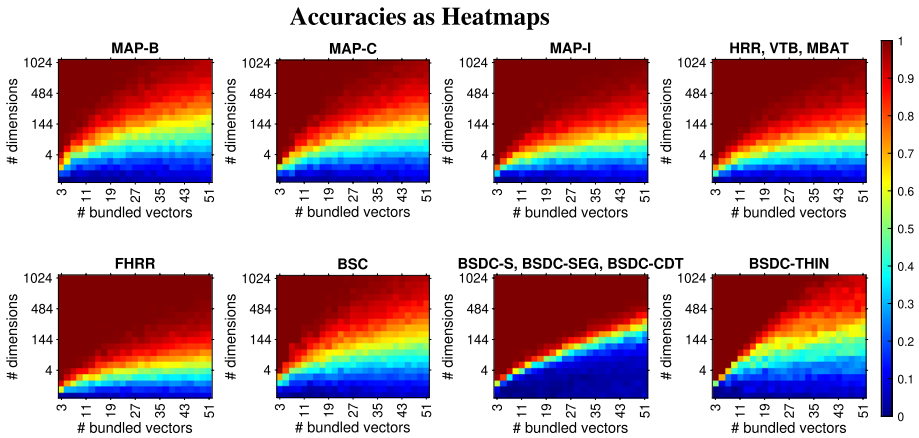


Fig. 3 Heat-maps showing the accuracies of different number of bundled vectors and numbers of dimensions

this is the result of the increased density after bundling without thinning. The last plot in Fig. 3 shows how the transition range between low and high accuracies increases when using an additional thinning (with maximum density 0.5)⁸.

For an easier access to the different VSAs performances in the capacity experiment, Fig. 4 summarizes the results of the heatmaps in 1-D curves. It provides an evaluation of the required number of dimensions to achieve almost perfect retrieval for different values of k . We selected a threshold of 99% accuracy, that means 99 of 100 query results are correct. A threshold of 100% would have been particularly sensitive to outliers, since a single wrong retrieval would prevent achieving the 100%, independent of the number of perfect retrieval cases. To make the comparison more accessible, we fit a straight line to the data points and plot the result as a dotted line.

Dense binary spaces need the highest number of dimensions, real-valued vectors a little less and the complex values require the smallest number of dimensions. As expected from the previous plots in Fig. 3, the binary sparse (**BSDC**, **BSDC-S**, **BSDC-SEG**) and the complex domain (**FHRR**) reach the most efficient results. They need fewer dimensions to bundle all vectors correctly. The sparse binary representations perform better than the dense binary vectors in this experiment. A more in-depth analysis of the general benefits of sparse distributed representations can be found in Ahmad and Scheinkman (2019). Particularly interesting is also the comparison between the **HRR** VSA from Plate (1995) and the complex-valued **FHRR** VSA from Plate (1994). Both the **FHRR** with the complex domain as well as the **HRR** architecture operate in a continuous space (where values in **FHRR** represent angles of unit-length complex numbers). However, operating with real values in a complex perspective increases the efficiency noticeably. Even if the **HRR** architecture is adapted to a range of $[-\pi, \pi]$ like the complex domain, the performance of the real VSA does not change remarkably. This is an interesting insight: If real numbers are treated as if they were angles of a complex number, then this increases the efficiency of bundling.

⁸ Since the BSDC architecture performance also depends on the given sparsity, we want to refer to Kleyko et al. (2018) for a more exhaustive sensitivity analysis of sparse vectors on a classification task.

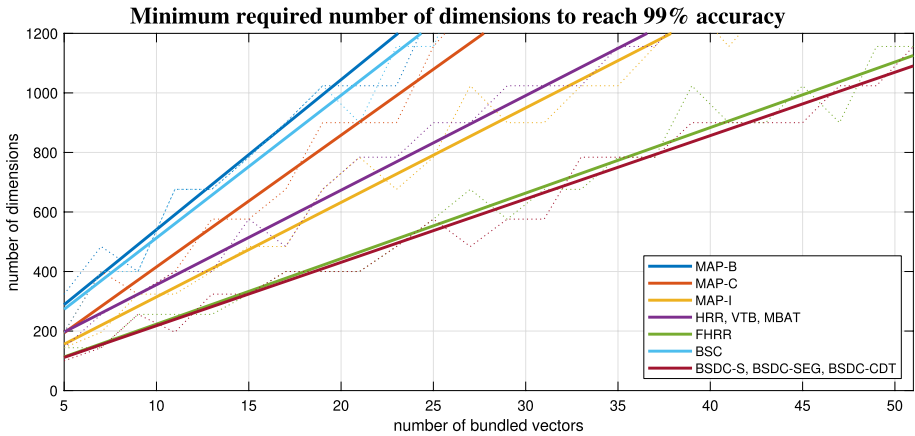


Fig. 4 Minimum required number of dimensions to reach 99% accuracy. The solid lines represent linear fitted curves. The flatter the curves/lines, the more efficient is the bundling. Keep in mind, different VSAs might have very different memory consumption per dimension

We want to emphasize again that different VSAs potentially require very different amounts of memory per dimension. Very interestingly, in these experiments, the sparse vectors require a low number of dimensions and are additionally expected to have particularly low memory consumption. A more in-depth evaluation of memory and computational demands is an important point for future work.

Besides the experimental evaluation of the bundle capacity, the literature provides analytical methods to predict the accuracy for a given number of bundled vectors and number of dimensions. Since it is not yet available for all of our evaluated VSAs, we have not used it in our comparison. However, we found a high accordance of our experimental results with the available analytical results. Further information about analytical capacity calculation can be found in Gallant and Okaywe (2013), Frady et al. (2018) and Kleyko (2018).

Influence of the item memory size In the above experiments, we used a fixed number of vectors in the item memory ($N = 1,000$). Plate (Plate 1994, p. 160 ff) describes a dependency between the size of the item memory and the accuracy of the superposition memory (bundled vectors) for Holographic Reduced Representations. The conclusion was that the number of vectors in the item memory (N) can be increased exponentially in the number of dimensions D while maintaining the retrieval accuracy. To evaluate the influence of the item memory size for all VSAs, we slightly modify our previous experimental setup. This time, we fix the number of bundled vectors to $k = 10$ and report the minimum number of dimensions that is required to achieve an accuracy of at least 99% for a varying number N of elements in the item memory.

The results can be seen in the Fig. 5 (using a logarithmic scale for the item memory size). Although the absolute performance varies between VSAs, the shape of the curves are in accordance with Plate's previous experiment on HRRs. Since there are no qualitative differences between the VSAs (the ordering of the graphs is consistent), our above comparison of VSAs for a varying number of bundled vectors k is presumably representative also for other item memory sizes N .

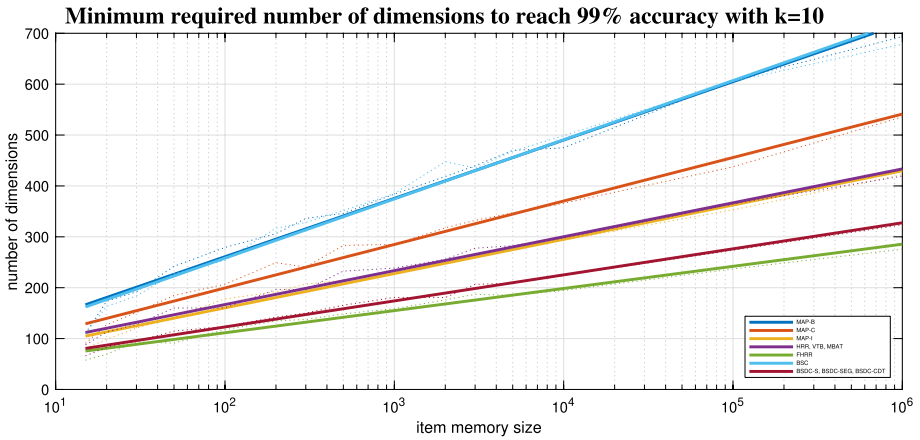


Fig. 5 Result of the capacity experiment with fixed number of neighbors and varying item memory size. Please note the logarithmic scale. The straight lines are fitted exponential functions

3.2 Performance of approximately invertible binding

The taxonomy in Fig. 1 includes three VSAs that only have an approximate inverse binding: **MAP-C**, **VTB** and **HRR**. The question is: *How good is the approximation of the binding inverse?* To evaluate the performance of the approximate inverses, we use a setup similar to Gosmann and Eliasmith (2019). We extended the experiment to compare the accuracy of approximate unbinding of the three relevant VSAs. The experiment is defined as follows: we start with an initial random vector \mathbf{v} and bind it sequentially with n other random vectors $\mathbf{r}_1 \dots \mathbf{r}_n$ to an encoded sequence \mathbf{S} (see Eq. 17). The task is to retrieve the elemental vector \mathbf{v} by sequentially unbinding the random vectors $\mathbf{r}_1 \dots \mathbf{r}_n$ from \mathbf{S} . The result is a vector \mathbf{v}' that should be highly similar to the original vector \mathbf{v} (see Eq. 18).

$$\mathbf{S} = ((\mathbf{v} \otimes \mathbf{r}_1) \otimes \mathbf{r}_2) \dots \otimes \mathbf{r}_n \tag{17}$$

$$\mathbf{v}' = \mathbf{r}_1 \oslash \dots (\mathbf{r}_{n-1} \oslash (\mathbf{r}_n \oslash \mathbf{S})) \tag{18}$$

We applied the described procedure for the 3 approximated VSAs (all exact-invertible bindings would produce 100% accuracy and are not shown in the plots) with $n = 40$ sequences and $D = 1024$ dimensions. The evaluation criterion is the similarity of \mathbf{v} and \mathbf{v}' , normalized to range [0, 1] (minimum to maximum possible similarity value). Results are shown in Fig. 6. In accordance with the results from Gosmann and Eliasmith (2019), the **VTB** binding and unbinding performs better than the circular convolution/correlation from **HRR**. It reaches the highest similarity over the whole range. The bind/unbind operator of the **MAP-C** architecture with values within the range $[-1, 1]$ performs slightly worse than **HRR**. In practice, VSA systems with such long sequences of approximate unbindings can incorporate a denoising mechanism. For example, a nearest neighbor search in an item memory with atomic vectors to clean up the resulting vector (often referred to as clean-up memory).

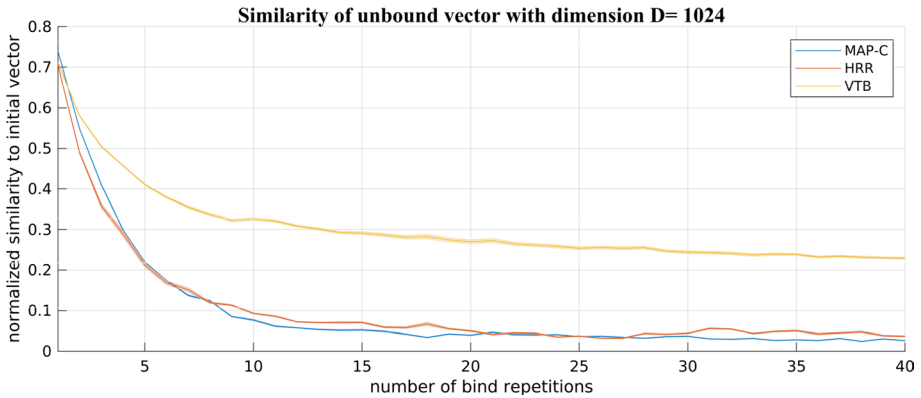


Fig. 6 Normalized similarity between the initial vector \mathbf{v} and the unbound sequence vector \mathbf{v}' with different numbers of sequences

3.3 Unbinding of bundled pairs

The third experiment combines the bundling, the binding and the unbinding operator in one scenario. It extends the example from the introduction, where we bundled three role-filler pairs to encode the knowledge about one country. A VSA allows querying for a filler by unbinding the role. Now, the question is: *How many property-value (role-filler) pairs can be bundled and still provide the correct answer to any query by unbinding a role?* This is similar to unbinding of a noisy representation and to the experiment on scaling properties of VSAs in (Eliasmith 2013, p. 141) but using only a single item memory size.

Similar to the bundle capacity experiment in the previous section 3.1, we create a database (item memory) of $N = 1,000$ random elemental vectors. We combine $2k$ (k roles and k fillers) randomly chosen elementary vectors from the item memory to k vector pairs by binding these two entities. The result are k bound pairs, equivalent to the property-value pairs from the USA example ($Name \otimes USA\dots$). These pairs are bundled to a single representation R (analog to the representation R_{USA}) which creates a noisy version of all bound pairs. The goal is to retrieve all $2k$ elemental vectors from the compact hypervector R by unbinding. The evaluation criterion is defined as follows: we compute the ratio (accuracy) of correctly recovered vectors to the number of all initial vectors ($2k$). As in the capacity experiment, we used a variable number of dimensions $D = 4\dots1156$ and a varying number of bundled pairs $k = 2\dots50$. Finally, we run the experiment 10 times and use the mean values.

Similar to the bundling capacity experiment (Sect. 3.1), we provide two plots: Fig. 7 presents the accuracies as heat-maps for all combinations of numbers of bundled pairs and dimensions, and Fig. 8 shows the minimum required number of dimensions to achieve 99% accuracy. Interestingly, the overall appearance of the heatmaps of the two BSDC architectures in Fig. 7 is roughly the same, but the **BSDC-SHIFT** has a noisy red area, which means that some retrievals failed even if the number of dimensions is high enough in general. The similar fuzziness can be seen at the heat-map of the **MBAT** VSA.

Again, Fig. 8 summarizes the results to 1-D curves. It contains more curves than in the previous section because some VSAs share the same bundling operator, but each has an individual binding operator. For example, the performance of the different **BSDC**

Accuracies as Heatmaps

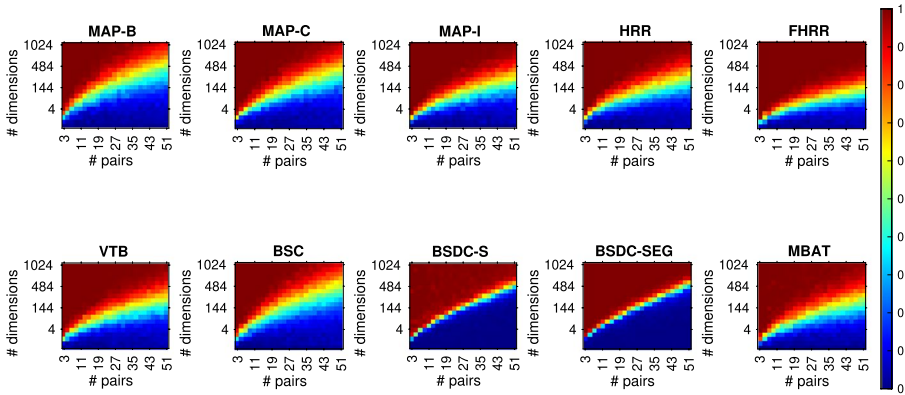


Fig. 7 Heat-maps showing the accuracies of different number of bundled vectors and numbers of dimensions

architectures varies. The sparse VSA with the segmental binding is more dimension-efficient than shifting the whole vector. However, all BSDC variants are less dimension-efficient than **FHRR** in this experiment, although they performed similar in the capacity experiment from Fig. 4. Furthermore, all VSAs based on the normal (Gaussian) distributed continuous space (**HRR**, **VTB** and **MBAT**) achieve very similar results. It seems that matrix binding (e.g. **MBAT** and **VTB**) does not significantly improve the binding and unbinding.

Finally, we evaluate the VSAs by comparing their accuracies to those of the capacity experiment from Sect. 3.1 as follows: We select the minimum required number of dimensions to retrieve either 15 bundled vectors (capacity experiment in Sect. 3.1) or 15 bundled pairs (bound vectors experiment). Table 2 summarizes the results and shows the increase between the bundle and the binding-plus-bundle experiment. Noticeably, there is

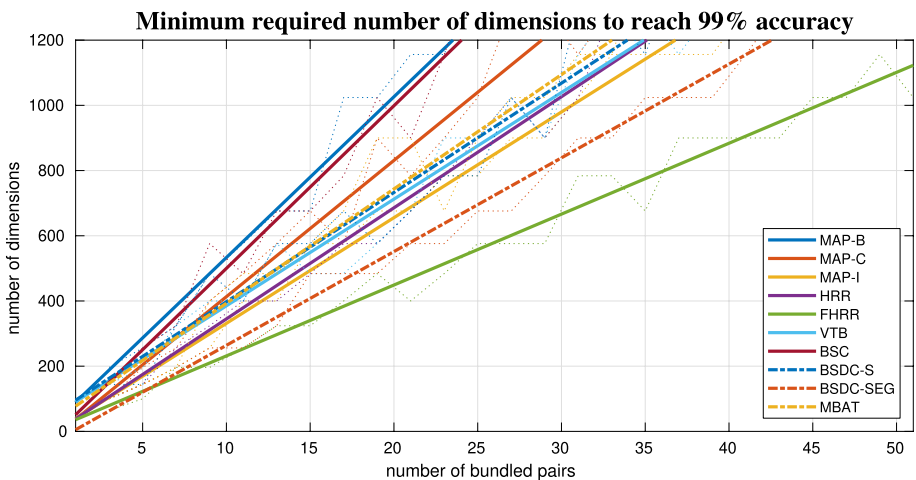


Fig. 8 Minimum required number of dimensions to reach 99% accuracy in unbinding of bundled pairs experiment. The solid lines represent linear fitted curves

Table 2 Comparison of the minimum required number of dimensions to reach a perfect retrieval of 15 bundled vectors and 15 bundled pairs (results are rounded to the tenth unit)

Vector space	# Dimensions to bundle 15 vectors	# Dimensions to bundle 15 pairs	Increase (%)
MAP-C	640	620	- 3
MAP-B	790	780	- 1
BSC	750	750	± 0
HRR	510	520	+ 2
FHRR	330	340	+ 3
MAP-I	470	490	+ 4
VTB	510	550	+ 7
MBAT	510	570	+ 11
BSDC-SEG	320	410	+ 22
BSDC-S	320	570	+ 44

Fourth column shows the growth between the first and the second experimental results (rounded to one unit)

a significant rise of the number of dimensions for the sparse binary VSA. It requires up to 44% larger vectors when using the bundling in combination with binding. However, the segmental shifting method with an increase of 22% works better than shifting the whole vector. One reason could be the increasing density during binding of sparsely distributed vectors because it uses only the disjunction without a thinning procedure. **MAP-C**, **MAP-B**, **MAP-I**, **HRR**, **FHRR** and **BSC** only show a marginal change of the required number of dimensions. Again, the complex **FHRR** VSA achieves the overall best performance regarding minimum number of dimensions and increase in order to account for pairs. However, this might result mainly from the good bundling performance rather than the better binding performance.

4 Practical applications

This section experimentally evaluates the different VSAs on two practical applications. The first is recognition of the language of a written text. The second is a task from mobile robotics: visual place recognition using real-world images, e.g., imagery of a 2800 km journey through Norway across different seasons. We chose these practical applications since the former is an established example from the VSA literature and the latter an example of a combination of VSAs with Deep Neural Networks. Again, we will compare VSA using the same number of dimensions. The actual memory consumption and computational cost per dimension can be quite different for each VSA. However, this will strongly depend on the available hard- and software.

4.1 Language recognition

For the first application, we selected a task that has previously been addressed using a VSA in the literature: recognizing the language of a written text. For instance, Joshi et al. (2017) presents a VSA approach to recognize the language of a given text from 21 possible languages. Each letter is represented by a randomly chosen hypervector (a vector symbolic

representation). To construct a meaningful representation of the whole language, short sequences of letters are combined in *n*-grams. The basic idea is to use VSA operations (binding, permutation, and bundling) to create the n-grams and compute an item memory vector for each language. The used permutation operator ρ is a simple shifting of the whole vector by a particular amount (e.g., permutation of order 5 is written as ρ^5). For example, the encoding of the word 'the' in a 3-gram (that combine exactly the three consecutive letters) is done as follows:

1. Basis is a fixed random hypervector for each letter: $\mathbf{v}_t, \mathbf{v}_h, \mathbf{v}_e$
2. The vector of each letter in the n-gram is permuted with the permutation operator according to the position in the n-gram: $\rho^0\mathbf{v}_t, \rho^1\mathbf{v}_h, \rho^2\mathbf{v}_e$
3. Permuted letter vectors are bound together to achieve a single vector that encodes the whole n-gram:

$$\mathbf{v}_{\text{the}} = \rho^0\mathbf{v}_t \otimes \rho^1\mathbf{v}_h \otimes \rho^2\mathbf{v}_e$$

The “learning” of a language is simply done by bundling all n-grams of a training dataset ($\mathbf{v}_{\text{english}} = \mathbf{v}_{\text{the}} + \mathbf{V}_{\dots}$). The result is a single vector representing the n-gram statistics of this language (i.e., the multiset of n-grams) and that can then be stored in an item memory. To later recognize the language of a given query text, the same procedure as for learning a language is repeated to obtain a single vector that represents all n-grams in the text, and a nearest neighbor query with all known language vectors in the item memory is performed.

We use the experimental setup from Joshi et al. (2017) with 21 languages and 3-grams to compare the performance of the different available VSAs. Since the matrix binding VSAs need a lot of time to learn the whole language vectors with our current implementation, we used a fraction of 1,000 training and 100 test sentences per language (which is 10% of the total dataset size from Joshi et al. (2017)).

Figure 9 shows the achieved accuracy of the different VSAs at the language recognition task for a varying number of dimensions between 100 and 2,000. In general, the more dimensions are used, the higher is the achieved accuracy. **MBAT**, **VTB** and **FHRR** need fewer dimensions to achieve high accuracy. It can be seen that the **VTB** binding is considerably better at this particular task than the original circular convolution binding of the **HRR** architecture (**HRR** is less efficient compared to **VTB**). Interestingly, the **FHRR** has almost the same accuracy as the architectures with matrix binding (**VTB** and **MBAT**) although it uses less costly element-wise operations for binding and bundling. Finally, **BSDC-CDT** was not evaluated on this task. Since it has no thinning process after bundling, bundling hundreds of n-gram vectors results in an almost completely filled vector which is unsuited for this task.

4.2 Place recognition

Visual place recognition is an important problem in the field of mobile robotics, e.g., it is an important means for loop closure detection in SLAM (Simulation Localization And Mapping). The following Sect. 4.2.1 will introduce this problem and outline the state-of-the-art approach SeqSLAM (Milford and Wyeth 2012). In Neubert et al. (2019b), we already described how a VSA can be used to encode the information from a sequence of images in a single hypervector and perform place recognition similarly to SeqSLAM. Approaching this problem with a VSA is particularly promising since the

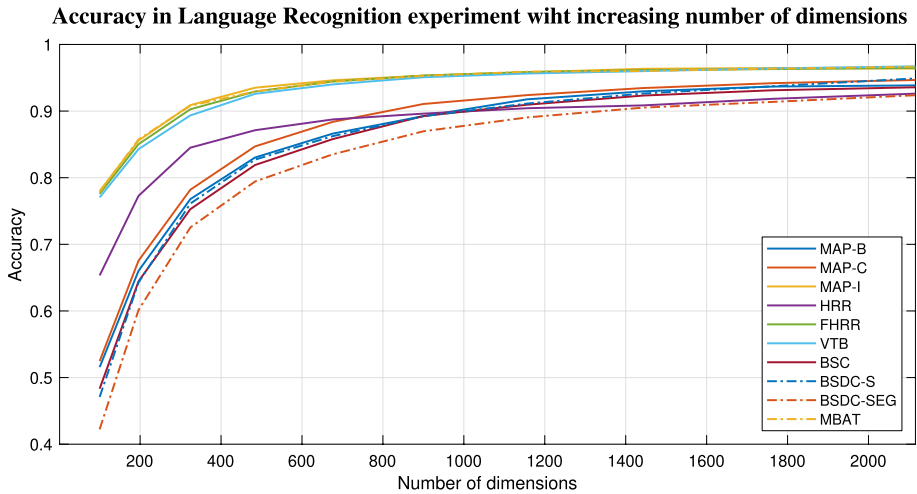


Fig. 9 Accuracy on the language recognition experiment with increasing number of dimensions. The results are smoothed with an average filtering with kernel size of three

image comparison is typically done based on the similarity of high-dimensional image descriptor vectors. The VSA approach has the advantage of only requiring a single vector comparison to decide about a matching—while SeqSLAM typically requires 5–10 times as many comparisons. After presentation of the CNN-based image encodings in Sects. 4.2.3, 4.2.4 will use this procedure from Neubert et al. (2019b) to evaluate the performance of the different VSAs.

4.2.1 Pairwise descriptor comparison and SeqSLAM

Place recognition is the problem of associating the robot's current camera view with one or multiple places from a database of images of known places (e.g., images of all previously visited locations). The essential source of information is a descriptor for each image that can be used to compute the similarity between each pair of a database and a query image. The result is a pairwise similarity matrix as illustrated on the left side of Fig. 11. The most similar pairs can then be treated as place matchings.

Place recognition is a special case of image retrieval. It differs from a general image retrieval task since the images typically have a temporal and spatial ordering—we can expect temporally neighbored images to show spatially neighbored places. A state-of-the-art place recognition method that exploits this additional constraint is SeqSLAM (Milford and Wyeth 2012), which evaluates short sequences of images in order to find correspondences between the query camera stream and the database images. Basically, SeqSLAM not only compares the current camera image to the database, but also the previous (and potentially the subsequent) images.

Algorithm 1 Simplified SeqSLAM core algorithm**Input:** Similarity matrix S of size $m \times n$, sequence length parameter d **Output:** New similarity matrix R

```

1: for  $i = 1 : m$  do
2:   for  $j = 1 : n$  do
3:     accSim = 0
4:     for  $k = -d : 1 : d$  do
5:       accSim +=  $S(i+k, j+k)$ 
6:     end for
7:      $R(i,j) = \text{accSim} / (2 \cdot d + 1)$ 
8:   end for
9: end for
10: return  $R$ 

```

Algorithm 1 illustrates the core processing of SeqSLAM in a simplified algorithmic listing. Input is a pairwise similarity matrix S . In order to exploit the sequential information, the algorithm iterates over all entries of S (the loops in lines 1 and 2). For each element the average similarities over the sequence of neighbored elements is computed in a third loop (line 4). This neighborhood sequence is illustrated as a red line in Fig. 11 (basically, this is a sparse convolution). This simple averaging is known to significantly improve the place recognition performance, in particular in case of changing environmental conditions (Milford and Wyeth 2012). The listing is intended to illustrate the core idea of SeqSLAM. It is simplified since border effects are ignored and since the original SeqSLAM evaluates different possible velocities (i.e. slopes of the neighborhood sequences). For more details, please refer to Milford and Wyeth (2012). The key benefit of the VSA approach to SeqSLAM is that it will allow to completely remove the inner-loop.

4.2.2 Evaluation procedure

To compare the performance of different place recognition approaches in our experiments, we use a standard evaluation procedure based on ground-truth information about place matchings (Neubert et al. 2019a). It is based on five datasets with available ground truth: **StLucia** Various Times of the Day (Glover et al. 2010), **Oxford RobotCar** (Maddern et al. 2017), **CMU Visual Localization** (Badino et al. 2011), **Nordland** (Sünderhauf et al. 2013) and **Gardens** Point Walking (Glover 2014). Given the output of a place recognition approach on a dataset (i.e., the initial matrix of pairwise similarities S or the output of SeqSLAM R), we run a series of thresholds on the similarities to get a set of binary matching decisions for each individual threshold. We use the ground truth to count true-positive (TP), false-positive (FP), and false-negative (FN) matchings, and further compute a point on the precision-recall curve for each threshold with precision $P = TP / (TP + FP)$ and recall $R = TP / (TP + FN)$. To obtain a single number that represents the place recognition performance, we report AUC, the area under the precision-recall curve (i.e., average precision, obtained by trapezoidal integration).

4.2.3 Encoding images for VSAs

Using VSAs in combination with real-world images for place recognition requires an image encoding into meaningful descriptors. Dependent on the particular vector space of the VSA the encoding will be different. We will first describe the underlying basic image descriptor, followed by an explanation and evaluation of the individual encodings for each VSA.

We use a basic descriptor similar to our previous work (Neubert et al. 2019a). Sünderhauf et al. (2015) showed that early convolutional layers of CNNs are a valuable source for creating robust image descriptors for place recognition. For example, the pre-trained AlexNet (Krizhevsky et al. 2012) generates the most robust image descriptors at the third convolution level. To use these as input for the place recognition pipeline, all images pass through the first three layers of AlexNet and the output tensor of size of $13 \times 13 \times 384$ is flattened to a vector of size 64,896. Next, we apply a dimension-wise standardization of the descriptors for each dataset following Schubert et al. (2020). Although this is already a high-dimensional vector, we use random projections in order to distribute information across dimensions and influence the number of dimensions: To obtain a N -dimensional vector (e.g. $N = 4,096$) from a M -dimensional space (e.g. $M = 64,896$), the original vector is multiplied by a random $M \times N$ matrix with values drawn from a Gaussian normal distribution. M is row-wise normalized. Such a dimensional reduction can lead to loss of information. The effect on the pairwise place recognition performance for each data set is shown in Fig. 10. It shows the AUC of pairwise comparison of both, the original descriptors and the dimension-reduced descriptors (calculated and evaluated as described in the section above). The plot supports that the random projection is a suitable method to reduce the dimensionality and distribute information, since the projected descriptors reach almost the same AUC as the original descriptors.

Afterwards the descriptors can be converted into the vector spaces of the individual VSAs (cf. table 1). Table 3 lists the encoding methods to convert the projected, standardized CNN descriptors to the different VSA vector spaces. It has to be noticed that the sLSBH method doubled the number of dimensions of the input vector (please refer to Neubert et al. (2019a) for details). The table also lists the influence of the encodings on the place recognition performance (mean and standard deviation of AUC change over all datasets). The performance change in the 4th column was computed by $(Acc_{projected} - Acc_{converted})/Acc_{projected}$.

It can be seen that the encoding method for **HRR**, **VTB** and **MBAT** VSAs does not influence the performance. In contrast, the conversion of the real-valued space into the sparse binary domain leads to significant performance losses (approx. 22%). However, this is mainly due to the fact that we compare the encoding of a *dense real* valued vector into a *sparse binary* vector of only twice the number of dimensions (a property of the used sLSBH procedure (Neubert et al. 2019a)). The encoding quality improves, if the number of dimensions in the sparse binary vector is increased. However, for consistency reasons, we keep the number of dimensions fixed. The density of the resulting sparse vectors is $1/\sqrt{2 \cdot D}$.

4.2.4 VSA SeqSLAM

The key idea of the VSA implementation of SeqSLAM is to replace the costly post-processing of the similarity matrix S in Algorithm 1 by a superposition of the information of neighbored images already in the high-dimensional descriptor vector of an image. Thus, the sequential information can be harnessed in a simple pairwise descriptor comparison and the inner-loop of SeqSLAM (line 4 in Algorithm 1) becomes obsolete.

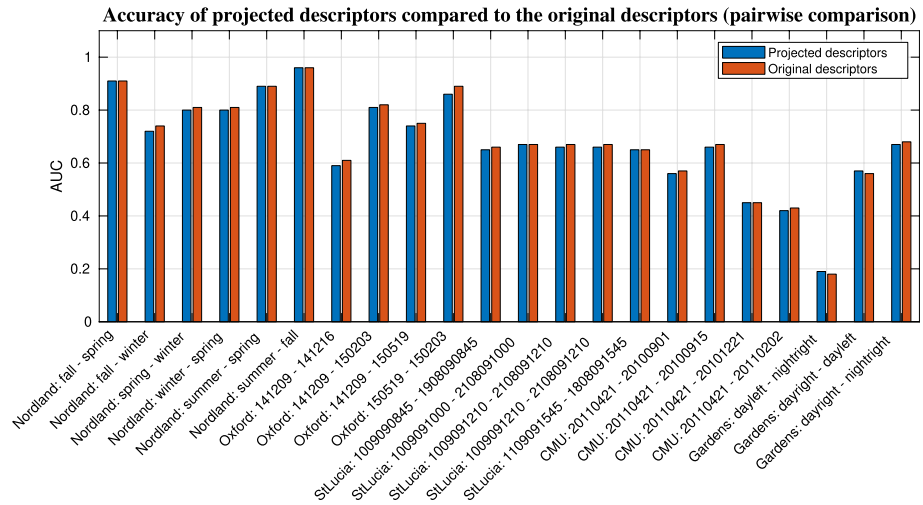


Fig. 10 AUC of original descriptors and projected descriptors (decrease number of dimensions) for each dataset. Evaluation based on pairwise comparison of database and query images

Table 3 Encoding methods

elements X of Space V	Encoding of input I	VSA	perf. change [%]
$X \in \{-1, 1\}^D$	$X = \begin{cases} 1 & I > 0 \\ -1 & I \leq 0 \end{cases}$	MAP-B	-2.2 ± 1.9
$X \in \{0, 1\}^D$	$X = \begin{cases} 1 & I > 0 \\ 0 & I \leq 0 \end{cases}$	BSC	-2.2 ± 1.9
$X \in [-1, 1]^D$	$X = \begin{cases} 1 & I \geq 1 \\ -1 & I \leq -1 \\ I & \text{else} \end{cases}$	MAP-C	-1 ± 1.3
$X \in \mathbb{R}^D$	$X = \frac{I}{\text{norm}(I)}$	HRR, VTB, MBAT	0
$X \in \mathbb{C}^D, X = e^{i\theta}$	$\theta = \arg(\mathcal{F}\{I\})$	FHRR	-0.9 ± 0.8
$X \in \{0, 1\}^{2 \cdot D}$	sLSBH Neubert et al. (2019b)	BSDC-S, BSDC-SEG	-21.9 ± 16

Last column represents the AUC change between the original data (after projection) and the converted data with pairwise comparison. The density of sLSBH is $1/\sqrt{2 \cdot D}$

This idea can be implemented as preprocessing of descriptors before the computation of the pairwise similarity matrix S . Each descriptor X_i in the database and query set is processed independently into a new descriptor vector Y_i that also encodes the neighboring descriptors:

$$Y_i = +_{k=-d}^d (X_{i+k} \otimes P_k) \tag{19}$$

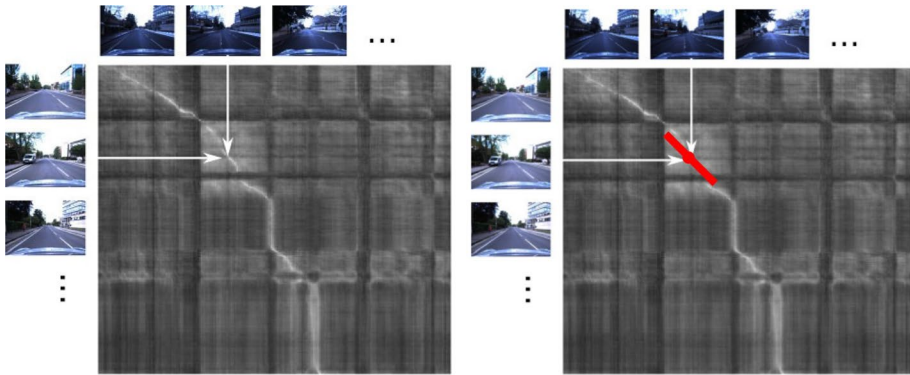


Fig. 11 Evaluation metric of the place recognition experiment. The gray tones images represent the similarity matrix (color encoded similarities between the database and query images – bright pixels corresponding to a high similarity). Left: pairwise comparison of database and query images. Right: sequence-based comparison of query and database images with the red line representing the sequence of compared images

Each image descriptor from the sequence neighborhood is bound to a static position vector P_k before bundling to encode the ordering of the images within the sequence. The position vectors are randomly chosen, but fixed across all database and query images. In a later pairwise comparison of two such vectors Y , only those descriptors X that are at corresponding positions within the sequence contribute to the overall similarity (due to the quasi-orthogonality of the random position vectors and the properties of the binding operator). In the following, we will evaluate the place recognition performance when implementing this approach with the different VSAs. Please refer to Neubert et al. (2019a) for more details on the approach itself.

4.2.5 Results

In the experiments, we use 4,096 dimensional vectors (except for sLSBH encodings with twice this number) and sequence length $d = 5$. Table 4 shows the results when using either the original SeqSLAM on an particular encoding or the VSA-implementation. The performance of the original SeqSLAM on the original descriptors (but with dimensionality reduction and standardization) can, e.g. be seen at the VTB column. To increase the readability, we highlighted the overall best results in bold and visualized the relative performance of a VSA to the corresponding original SeqSLAM with colored arrows. In most cases, the VSA approaches can approximate the SeqSLAM method with essentially the same AUC. Particularly the real-valued vector spaces (**MAP-C**, **HRR**, **VTB**) yield good AUC in both the encoding itself (Table 3) and the sequence-based place recognition task. **MAP-C** achieves 100% AUC on the Nordland dataset (which is even slightly better than the SeqSLAM algorithm) and has no considerable AUC reduction in any other datasets. Also the **VTB** and **MBAT** architectures achieve very similar results to the original SeqSLAM approach. However, it has to be noticed that these VSAs use matrix binding methods, which leads to a high computational effort compared to element-wise binding operations. The performance of the sparse VSAs (**BSDC-S**, **BSDC-SEG**) varies, including cases where the performance is considerably worse than the original SeqSLAM (which in turn achieves surprisingly good results given the overall performance drop of the sparse encoding from Table 3).

Table 4 Results (AUC) of the place recognition experiment with original datasets

Dataset	Database	Query	MAP-B		MAP-C		MAP-I		HRR		VTB		MBAT		FHRR		BSC		BSDC-S		BSDC-SEG		
			orig	VSA	orig	VSA	orig	VSA	orig	VSA	orig	VSA	orig	VSA	orig	VSA	orig	VSA	orig	VSA	orig	VSA	
Nordland	fall	spring	0.98	1.00	0.98	1.00	0.98	1.00	0.98	1.00	0.98	1.00	0.98	1.00	0.98	1.00	0.98	1.00	0.92	0.97	0.92	0.97	
	fall	winter	0.98	0.99	0.99	1.00	0.98	0.99	0.99	1.00	0.99	1.00	0.99	1.00	0.98	0.98	0.89	0.87	0.89	0.87	0.89	0.85	
	spring	winter	0.96	0.97	0.96	0.99	0.96	0.99	0.96	0.98	0.96	0.98	0.96	0.99	0.96	0.99	0.96	0.97	0.82	0.83	0.82	0.84	
	winter	spring	0.96	0.97	0.96	0.99	0.96	0.98	0.96	0.98	0.96	0.98	0.96	0.99	0.96	0.97	0.84	0.83	0.84	0.83	0.84	0.95	
Nordland	spring	spring	0.98	0.99	0.98	1.00	0.98	1.00	0.98	1.00	0.98	1.00	0.98	1.00	0.98	0.99	0.98	1.00	0.97	0.97	0.93	0.97	
	summer	spring	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.97	0.99	0.97	0.99	
	summer	summer	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.67	0.53	0.40	0.36	
	fall	fall	0.67	0.52	0.66	0.64	0.67	0.67	0.67	0.67	0.63	0.60	0.63	0.61	0.63	0.53	0.63	0.53	0.40	0.36	0.40	0.36	
Oxford	141209	141216	0.85	0.79	0.85	0.82	0.85	0.81	0.84	0.85	0.84	0.84	0.80	0.84	0.81	0.84	0.80	0.85	0.80	0.65	0.55	0.65	0.61
Oxford	141209	150519	0.82	0.82	0.81	0.83	0.82	0.75	0.80	0.81	0.80	0.76	0.81	0.83	0.82	0.80	0.83	0.81	0.83	0.81	0.71	0.60	0.60
Oxford	150519	150203	0.91	0.89	0.91	0.90	0.91	0.88	0.91	0.91	0.91	0.89	0.91	0.91	0.91	0.89	0.91	0.89	0.84	0.75	0.84	0.67	0.67
StLucia	100909-0845	190809-0845	0.83	0.78	0.83	0.83	0.83	0.81	0.83	0.83	0.83	0.82	0.83	0.82	0.83	0.81	0.83	0.79	0.83	0.77	0.85	0.80	0.80
StLucia	100909-1000	210809-1000	0.87	0.82	0.88	0.86	0.87	0.85	0.88	0.86	0.88	0.86	0.88	0.86	0.87	0.85	0.87	0.82	0.86	0.79	0.87	0.82	0.82
StLucia	100909-1210	210809-1210	0.89	0.79	0.88	0.84	0.89	0.83	0.89	0.84	0.89	0.84	0.89	0.84	0.89	0.84	0.89	0.80	0.88	0.76	0.88	0.77	0.77
StLucia	100909-1210	210809-1210	0.89	0.79	0.88	0.84	0.89	0.83	0.89	0.84	0.89	0.84	0.89	0.84	0.89	0.84	0.89	0.80	0.88	0.76	0.88	0.77	0.77
StLucia	110909-1545	180809-1545	0.85	0.76	0.85	0.82	0.85	0.80	0.85	0.80	0.85	0.81	0.85	0.82	0.85	0.80	0.85	0.76	0.84	0.75	0.84	0.76	0.76
CNU	20110421	20100901	0.64	0.65	0.63	0.64	0.64	0.60	0.63	0.62	0.63	0.61	0.63	0.61	0.63	0.65	0.64	0.65	0.57	0.50	0.57	0.49	0.49
CNU	20110421	20100915	0.75	0.71	0.74	0.74	0.75	0.70	0.74	0.73	0.74	0.73	0.74	0.72	0.74	0.73	0.75	0.72	0.66	0.58	0.66	0.56	0.56
CNU	20110421	20101221	0.56	0.54	0.55	0.54	0.55	0.53	0.55	0.54	0.55	0.53	0.55	0.56	0.56	0.56	0.55	0.55	0.31	0.22	0.31	0.20	0.20
CNU	20110421	20110202	0.52	0.45	0.51	0.50	0.52	0.47	0.51	0.48	0.51	0.49	0.51	0.48	0.51	0.48	0.52	0.46	0.50	0.37	0.50	0.37	0.37
Gardens	day-left	night-right	0.27	0.19	0.26	0.25	0.27	0.22	0.28	0.30	0.28	0.26	0.28	0.27	0.29	0.22	0.27	0.22	0.27	0.22	0.31	0.13	0.31
Gardens	day-right	day-left	0.77	0.66	0.77	0.74	0.77	0.67	0.77	0.75	0.77	0.72	0.77	0.73	0.76	0.72	0.77	0.67	0.71	0.55	0.71	0.55	0.71
Gardens	day-right	night-right	0.80	0.74	0.80	0.79	0.80	0.78	0.81	0.79	0.81	0.80	0.81	0.79	0.82	0.79	0.80	0.73	0.78	0.64	0.78	0.63	0.63

Sequence length is 5 and all sequence methods use constant velocity. The colored arrows indicate large ($\geq 25\%$), medium ($\geq 10\%$), or no ($< 10\%$) deviation from SeqSLAM (orig.)

5 Summary and conclusion

We discussed and evaluated available VSA implementations theoretically and experimentally. We created a general overview of the most important properties and provided insights especially to the various implemented binding operators (taxonomy of Fig. 1). It was shown that self-inverse binding operations benefit in applications such as analogical reasoning (“What is the Dollar of Mexico?”). On the other hand, these self-inverse architectures, like **MAP-B** and **MAP-C**, show a trade-off between an exactly working binding (by using a binary vectors space like $\{0, 1\}$ or $\{-1, 1\}$) or a high bundling capacity (by using real-valued vectors). In the bundling capacity experiment, the sparse binary VSA **BSDC** performed well and required only a small number of dimensions. However, in combination with binding, the required number of dimensions increased significantly (and also including the thinning procedure did not improve this result). Regarding the real-world application to place recognition, the sparse VSAs did not perform as well as other VSAs. Presumably, this can be improved by a different encoding approach or by using a higher number of dimensions (which would be feasible given the storage efficiency of sparse representations). High performance at both synthetic and real-world experiments could be observed in the simplified complex architecture **FHRR** that uses only the angles of the complex values. Since this architecture is not self-inverse, it requires a separate unbinding operation and cannot solve the “What is the dollar of Mexico?” example by Kanerva’s elegant approach. However, it could presumably be solved using other methods that iteratively process the knowledge tree (e.g., the readout machine in Plate (1995)), but come at increased computational costs. Furthermore, the two matrix binding VSAs (**MABT** and **VTB**) also show good results in the practical applications of language and place recognition. However, the drawback of these architecture is the high computational effort for binding.

This paper, in particular the taxonomy of binding operations, revealed a very large diversity in available VSAs and the necessity of continued efforts to systematize these approaches. However, the theoretical insights from this paper together with the provided experimental results on synthetic and real data can be used to select an appropriate VSA for new applications. Further, they are hopefully also useful for the development of new VSAs.

Although the memory consumption and computational costs per dimension can significantly vary between VSAs, the experimental evaluation compared different VSAs using a common number of dimensions. We made this decision since the actual costs depend on several factors like the underlying hard- and software, or the required computational precision for the current task. For example, some high-level languages like Matlab do not well support binary representations and not all CPUs support half-precision floats. We consider the number of dimensions as an intuitive common basis for comparison between VSAs that can later be converted to memory consumption and computational costs once the influencing factors for a particular application are clear. Recent in-memory implementations of VSA operators (Karunaratne et al. 2020) are important steps towards VSA specific hardware. Nevertheless, a more in-depth evaluation of resource consumption of the different VSAs is a very important part of future work. However, this will require additional design decisions and assumptions about properties of the underlying hard- and software.

Finally, we want to repeat the importance of permutations for VSAs. However, as explained in Sect. 2, we decided to not particularly evaluate differences in combination with permutations since they are applied very similarly in all VSAs (however, simple permutations were used in the language recognition task).

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ahmad S, Hawkins J (2015) Properties of sparse distributed representations and their application to hierarchical temporal memory. *CoRR*
- Ahmad S, Scheinkman L (2019) How can we be so dense? The benefits of using highly sparse representations. *CoRR*
- Badino H, Huber D, Kanade T (2011) Visual topometric localization. In: Proceedings of the intelligent vehicles symposium
- Bellman RE (1961) Adaptive control processes: a guided tour. MIT Press, Cambridge
- Beyer K, Goldstein J, Ramakrishnan R, Shaft U (1999) When is “nearest neighbor” meaningful? In: Database theory—ICDT99. Springer, Berlin, pp 217–235
- Cheung B, Terekhov A, Chen Y, Agrawal P, Olshausen B (2019) Superposition of many models into one. In: Advances in neural information processing systems 32. Curran Associates, Inc, pp 10868–10877
- Danihelka I, Wayne G, Uria B, Kalchbrenner N, Graves A (2016) Associative long short-term memory. In: Proceedings of the 33rd international conference on machine learning, vol 48. PMLR, New York, USA, pp 1986–1994
- Eliasmith C (2013) How to build a brain: a neural architecture for biological cognition. Oxford University Press, Oxford
- Fraday EP, Kleyko D, Sommer FT (2021) Variable binding for sparse distributed representations: theory and applications. *IEEE Trans Neural Netw Learn Syst*, pp 1–14. <https://doi.org/10.1109/TNNLS.2021.3105949>. <https://ieeexplore.ieee.org/document/9528907>
- Fraday EP, Kleyko D, Sommer FT (2018) A theory of sequence indexing and working memory in recurrent neural networks. *Neural Comput* 30(6):1449–1513. <https://doi.org/10.1162/NECO>
- Gallant SI, Okaywe TW (2013) Representing objects, relations, and sequences. *Neural Comput* 25:2038–2078
- Gayler RW (1998) Multiplicative binding, representation operators, and analogy. In: Advances in analogy research: integration of theory and data from the cognitive, computational, and neural sciences. New Bulgarian University
- Gayler RW (2003) Vector symbolic architectures answer Jackendoff's challenges for cognitive neuroscience. In: Proceedings of the ICCS/ASCS international conference on cognitive science, pp 133–138. Sydney, Australia
- Gayler RW, Levy SD (2009) A distributed basis for analogical mapping. *New Frontiers in Analogy Research, Proceedings of the second international conference on analogy, ANALOGY-2009*, pp 165–174
- Glover A (2014) Day and night with lateral pose change datasets. <https://wiki.qut.edu.au/display/cyphy/Day+and+Night+with+Lateral+Pose+Change+Datasets>
- Glover A, Maddern W, Milford M, Wyeth G (2010) FAB-MAP + RatSLAM: appearance-based SLAM for multiple times of day. In: Proceedings of the international conference on robotics and automation
- Gosmann J, Eliasmith C (2019) Vector-derived transformation binding: an improved binding operation for deep symbol-like processing in neural networks. *Neural Comput* 31:849–869
- Joshi A, Halseth JT, Kanerva P (2017) Language geometry using random indexing. *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)* 10106 LNCS:265–274. https://doi.org/10.1007/978-3-319-52289-0_21
- Kanerva P (2010) What we mean when we say what the Dollar of Mexico? Prototypes and mapping in concept space. In: AAAI fall symposium: quantum informatics for cognitive, social, and semantic processes, pp 2–6

- Kanerva P (1996) Binary spatter-coding of ordered K-tuples. *Artif Neural Netw ICANN Proc* 1112:869–873
- Kanerva P (2009) Hyperdimensional computing: an introduction to computing in distributed representation with high-dimensional random vectors. *Cogn Comput* 1(2):139–159
- Kanerva P, Sjoedin G, Kristoferson J, Karlsson R, Levin B, Holst A, Karlgren J, Sahlgren M (2001) Computing with large random patterns. <http://eprints.sics.se/3138/%5Cnhttp://www.rni.org/kanerva/rwics.pdf>
- Karunaratne G, Le Gallo M, Cherubini G, Benini L, Rahimi A, Sebastian A (2020) In-memory hyperdimensional computing. *Nat Electron* 3(6):327–337. <https://doi.org/10.1038/s41928-020-0410-3>
- Karunaratne G, Schmuck M, Le Gallo M, Cherubini G, Benini L, Sebastian A, Rahimi A (2021) Robust high-dimensional memory-augmented neural networks. *Nat Commun* 12(1):1–12. <https://doi.org/10.1038/s41467-021-22364-0>
- Kelly MA, Blostein D, Mewhort DJ (2013) Encoding structure in holographic reduced representations. *Can J Exp Psychol* 67(2):79–93. <https://doi.org/10.1037/a0030301>
- Kleyko D (2018) Vector symbolic architectures and their applications. Ph.D. thesis, Luleå University of Technology, Luleå, Sweden
- Kleyko D, Osipov E, Gayler RW, Khan AI, Dyer AG (2015) Imitation of honey bees concept learning processes using vector symbolic architectures. *Biol Inspired Cogn Archit* 14:57–72. <https://doi.org/10.1016/j.bica.2015.09.002>
- Kleyko D, Rahimi A, Rachkovskij DA, Osipov E, Rabaey JM (2018) Classification and recall with binary hyperdimensional computing: tradeoffs in choice of density and mapping characteristics. *IEEE Trans Neural Netw Learn Syst* 29(12):5880–5898. <https://doi.org/10.1109/TNNLS.2018.2814400>
- Kleyko D, Rahimi A, Gayler RW, Osipov E (2020) Autoscaling Bloom filter: controlling trade-off between true and false positives. *Neural Comput Appl* 32(8):3675–3684. <https://doi.org/10.1007/s00521-019-04397-1>
- Kleyko D, Osipov E, Papakonstantinou N, Vyatkin V, Mousavi A (2015) Fault detection in the hyperspace: towards intelligent automation systems. In: 2015 IEEE 13th international conference on industrial informatics (INDIN), pp 1219–1224. <https://doi.org/10.1109/INDIN.2015.7281909>
- Kleyko D, Rahimi A, Rachkovskij DA, Osipov E, Rabaey JM (2018) Classification and recall with binary hyperdimensional computing: tradeoffs in choice of density and mapping characteristics. *IEEE Trans Neural Netw Learn Syst*, pp 1–19. <https://doi.org/10.1109/TNNLS.2018.2814400>
- Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems* 25, pp 1097–1105. Curran Associates, Inc
- Laiho M, Poikonen JH, Kanerva P, Lehtonen E (2015) High-dimensional computing with sparse vectors. In: *IEEE biomedical circuits and systems conference: engineering for healthy minds and able bodies, BioCAS 2015—proceedings*, pp 1–4. IEEE. <https://doi.org/10.1109/BioCAS.2015.7348414>
- Maddern W, Pascoe G, Linegar C, Newman P (2017) 1 Year, 1000km: the Oxford RobotCar dataset. *Int J Robot Res* 36(1):3–15. <https://doi.org/10.1177/0278364916679498>
- Milford M, Wyeth GF (2012) Seqslam: visual route-based navigation for sunny summer days and stormy winter nights. In: *Proceedings of the IEEE international conference on robotics and automation (ICRA)*
- Neubert P, Schubert S (2021) Hyperdimensional computing as a framework for systematic aggregation of image descriptors. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp 16938–16947. <https://doi.org/10.1109/CVPR46437.2021.01666>
- Neubert P, Schubert S, Protzel P (2019a) A neurologically inspired sequence processing model for mobile robot place recognition. *IEEE Robot Autom Lett* 4(4):3200–3207. <https://doi.org/10.1109/LRA.2019.2927096>
- Neubert P, Schubert S, Protzel P (2019b) An introduction to high dimensional computing for robotics. In: *German journal of artificial intelligence special issue: reintegrating artificial intelligence and robotics*. Springer
- Neubert P, Schubert S, Schlegel K, Protzel P (2021) Vector semantic representations as descriptors for visual place recognition. In: *Proceedings of robotics: science and systems (RSS)*. <https://doi.org/10.15607/RSS.2021.XVII.083>
- Osipov E, Kleyko D, Legalov A (2017) Associative synthesis of finite state automata model of a controlled object with hyperdimensional computing. In: *IECON 2017-43rd annual conference of the IEEE industrial electronics society*, pp 3276–3281. <https://doi.org/10.1109/IECON.2017.8216554>
- Plate TA (1994) Distributed representations and nested compositional structure. Ph.D. thesis, University of Toronto, Toronto, Ont., Canada, Canada
- Plate TA (1997) A common framework for distributed representation schemes for compositional structure. In: *Connectionist systems for knowledge representations and deduction* (July), 15–34

- Plate TA (1995) Holographic reduced representations. *IEEE Trans Neural Netw* 6(3):623–641. <https://doi.org/10.1109/72.377968>
- Plate TA (2003) Holographic reduced representation: distributed representation for cognitive structures. CSLI Publications, New York
- Rachkovskij DA (2001) Representation and processing of structures with binary sparse distributed codes. *IEEE Trans Knowl Data Eng* 13(2):261–276. <https://doi.org/10.1109/69.917565>
- Rachkovskij DA, Kussul EM (2001) Binding and normalization of binary sparse distributed representations by context-dependent thinning. *Neural Comput* 13(2):411–452. <https://doi.org/10.1162/089976601300014592>
- Rachkovskij DA, Slipchenko SV (2012) Similarity-based retrieval with structure-sensitive sparse binary distributed representations. *Comput Intell* 28(1):106–129. <https://doi.org/10.1111/j.1467-8640.2011.00423.x>
- Rahimi A, Datta S, Kleyko D, Frady EP, Olshausen B, Kanerva P, Rabaey JM (2017) High-dimensional computing as a nanoscalable paradigm. *IEEE Trans Circuits Syst I Regul Pap* 64(9):2508–2521. <https://doi.org/10.1109/TCSI.2017.2705051>
- Schubert S, Neubert P, Protzel P (2020) Unsupervised learning methods for visual place recognition in discretely and continuously changing environments. In: International conference on robotics and automation (ICRA)
- Smolensky P (1990) Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artif Intell* 46(1–2):159–216
- Sünderhauf N, Neubert P, Protzel P (2013) Are we there yet? challenging seqslam on a 3000 km journey across all four seasons. In: Proceedings of the workshop on long-term autonomy at the international conference on robotics and automation
- Sünderhauf N, Shirazi S, Dayoub F, Upcroft B, Milford M (2015) On the performance of ConvNet features for place recognition. In: IEEE international conference on intelligent robots and systems, pp 4297–4304. <https://doi.org/10.1109/IROS.2015.7353986>
- Thrun S, Burgard W, Fox D (2005) Probabilistic robotics (intelligent robotics and autonomous agents). The MIT Press, Cambridge
- Tissera MD, McDonnell MD (2014) Enabling question answering in the MBAT vector symbolic architecture by exploiting orthogonal random matrices. In: Proceedings—2014 IEEE international conference on semantic computing, ICSC 2014, pp 171–174. <https://doi.org/10.1109/ICSC.2014.38>
- Widdows D (2004) Geometry and Meaning. Center for the Study of Language and Information Stanford, CA
- Widdows D, Cohen T (2015) Reasoning with vectors: a continuous model for fast robust inference. *Logic J IGPL Interest Group Pure Appl Log* 2:141–173
- Yerxa T, Anderson A, Weiss E (2018) The hyperdimensional stack machine. In: Poster at cognitive computing
- Yilmaz O (2015) Symbolic computation using cellular automata-based hyperdimensional computing. *Neural Comput* 27(12):2661–2692. https://doi.org/10.1162/NECO_a_00787

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.